

密级：_____

浙江大学

硕 士 学 位 论 文



论文题目 模糊重构算法及其分类应用的研究

作者姓名 代成超

指导教师 汤永川

学科(专业) 计算机应用技术

所在学院 计算机科学与技术学院

提交日期 2016-03-08

A Dissertation Submitted to Zhejiang
University for the Degree of
Master of Engineering



TITLE: Fuzzy Reconstruction Algorithm
and its Application about Classification

Author: Chengchao Dai

Supervisor: Yongchuan Tang

Subject: Computer Application Technology

College: Computer Science and Technology

Submitted Date: 2016-03-08

浙江大学研究生学位论文独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 浙江大学 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：

签字日期：

年 月 日

学位论文版权使用授权书

本学位论文作者完全了解 浙江大学 有权保留并向国家有关部门或机构送交本论文的复印件和磁盘，允许论文被查阅和借阅。本人授权 浙江大学 可以将学位论文的全部或部分内容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本授权书）

学位论文作者签名：

导师签名：

签字日期： 年 月 日

签字日期： 年 月 日

摘要

模糊隶属函数是模糊集中的一个重要概念，可以用来表示概念的语义信息。根据问题的性质和专家经验定义模糊隶属函数，该方法具有较强的主观性，可扩展性不强，适用于低维的小数据集。但不适用于高维的大数据。模糊隶属函数的确定需要客观地反映事物的本质，为此，本文提出一种新颖的模糊重构算法，从数据中学习隶属函数，用于构建模糊推理系统与大数据之间的桥梁。而且该方法可以获得重构结果，可解释性较强。测试样本在给定不同类别的训练样本上重构，根据重构系数获得模糊隶属度。测试样本在训练样本上的重构系数比较大，也即表明测试样本和这些训练样本属于同一类的可能性比较大。根据最大隶属度原则进行分类。

由于字典过完备，模糊重构的优化速度过慢。通过分析重构系数，发现只有部分样本参与重构，进而我们提出了基于近邻选择的模糊重构算法，选择和测试样本较为相似的部分训练样本作为字典，减少字典数目，由此可以提高优化速度。基于近邻选择的模糊重构能够快速获得模糊隶属函数，可以看作是一般化的 KNN 算法，但比 KNN 具有较强的鲁棒性，而且 K 值的选择较为简单。基于近邻选择的模糊重构较好地平衡了分类准确率和运行速度之间的矛盾。在一些常用的人脸、手写数字等大型标准数据集上进行了分类实验，该算法和稀疏表示 SRC 的分类效果相当，但是始终优于保序模糊重构、KNN 和模糊 KNN。

关键词：模糊隶属函数，数据降维，模糊重构，保序模糊重构，基于近邻选择的模糊重构，稀疏表示，人脸识别，手写数字识别

Abstract

In fuzzy set theory, membership function is very important and it is a tool to represent the semantic information of concepts. Usually, membership functions are defined based on the analysis of specific problems and the experience of experts. Such definitions can be subjective and not scalable, and thus can only be applied to small data sets of low dimensions instead of large ones or those data of high dimensions. The definition of membership function is expected to characterize the relation between concepts and entities objectively. Therefore, a novel fuzzy reconstruction algorithm (FR) is proposed in this paper, which can learn membership functions from data and build the bridge between fuzzy reasoning system and big data. One of the advantage this method boasts is that it can give interpretable results from reconstruction. Given a test sample, the method performs constrained linear reconstruction of this sample on training samples of different classes, and based on the solution of this reconstruction fuzzy membership can be obtained. The membership is proportional to the reconstruction coefficients; the larger reconstruction coefficients training samples of a particular class gains, the higher membership the test sample to this class is. Accordingly, the test sample can be classified to the class with the highest membership.

The main challenge in application of this method is that computation in reconstruction process can be intractable when we have over-complete training data set. However, it is observed that only a small part of samples contributes to the reconstruction in most cases. Therefore, we further develop the fuzzy reconstruction method based on K-Nearest-Neighbor (KNN) algorithm, which only consider those training samples similar to the test sample in reconstruction. This strategy can significantly reduce computation and speed up the optimization. From another aspect, this improved method can be regarded as a generalized KNN algorithm with stronger robustness but less sensitivity to the choice of parameter K . This method balances the contradiction between the accuracy and the optimization speed. Experiments on the

publicly available face datasets and digit datasets demonstrate FR algorithms based on KNN are comparable with SRC and outperform order preserving fuzzy reconstruction, KNN and Fuzzy KNN.

Keywords: Fuzzy Membership Function, Dimensionality Reduction, Fuzzy Reconstruction, Order Preserving Fuzzy Reconstruction ,Fuzzy Reconstruction Based on KNN ,Sparse Representation, Face Recognition, Digit Recognition

目录

摘要	i
Abstract.....	ii
图目录	III
表目录	IV
第 1 章 绪论	1
1.1 课题背景	1
1.2 国内外研究现状	3
1.3 本文的主要研究内容与贡献	6
1.4 论文组织结构	8
1.5 本章小结	9
第 2 章 数据降维方法	10
2.1 维灾难	10
2.2 无监督的降维方法	13
2.2.1 主成分分析(PCA).....	13
2.2.2 非负矩阵分解(NMF).....	14
2.2.3 等测地距离映射(Isomap).....	16
2.2.4 局部线性嵌入(LLE)	17
2.2.5 保局部投影(LPP).....	18
2.3 有监督的降维方法	19
2.3.1 Fisher 线性判别分析(LDA)	19
2.4 降维效果的评估	21
2.4.1 可视化效果	22
2.4.2 分类效果	24
2.5 本章小结	27
第 3 章 模糊重构算法	28
3.1 K 近邻算法和模糊 K 近邻算法	28
3.1.1 K 近邻算法(KNN).....	28
3.1.2 模糊 K 近邻算法(FKNN).....	29
3.2 稀疏表示 SRC.....	29
3.3 模糊重构	30
3.4 保序模糊重构	32
3.5 基于近邻选择的模糊重构	34
3.6 本章小结	35
第 4 章 模糊重构算法的分类实验	37
4.1 数据集简介	37
4.2 FR、KFR、KNN 和 FKNN 对比实验	38

4.3 基于近邻选择的各类算法对比实验	40
4.4 重构系数分析	44
4.5 调整重构系数改善分类性能	46
4.6 本章小结	47
第 5 章 总结和展望	49
5.1 全文总结	49
5.2 未来的研究工作展望	50
参考文献	52
致谢	57

图目录

图 1.1 模式分类的一般过程	3
图 2.1 过拟合现象	12
图 2.2 PCA 投影过程说明	13
图 2.3 NMF 学习了部分脸表示, PCA 学习的是人脸的整体表示	15
图 2.4 Isomap 的处理过程	17
图 2.5 LLE 降维过程	18
图 2.6 PCA 和 LDA 投影效果对比	20
图 3.1 模糊重构算法的整体过程	31
图 3.2 测试样本相对 $K(K=38)$ 类的隶属度	32
图 3.3 KFR 和 KNN 的对比	35
图 4.1 FR、KFR、KNN 以及 FKNN 在 Extend_Yale_B 人脸数据集上的分类准确率对比	39
图 4.2 FR、KFR、KNN 以及 FKNN 在不同类别的样本上的分类准确率	40
图 4.3 Yale 数据集, 随着 K 取值的变化五种算法的分类效果对比	41
图 4.4 PIE 数据集, 随着 K 取值的变化五种算法的分类效果对比	42
图 4.5 Pendigits 数据集, 随着 K 取值的变化五种算法的分类效果对比	43
图 4.6 USPS 数据集, 随着 K 取值的变化五种算法的分类效果对比	43
图 4.7 Sonar 数据集, 随着 K 取值的变化五种算法的分类效果对比	44
图 4.8 测试样本利用不同的隶属度强度作为分类标准时被正确分类的样本数分布情况	45
图 4.9 不同的分类标准下, 被正确分类的样本相对各类隶属度平均分布情况	46

表目录

表 2.1 数据降维后, KNN 分类效果对比.....	25
表 2.2 数据降维后, 基于近邻选择的模糊重构分类效果对比	26
表 4.1 数据集简介	38
表 4.2 选择不同的阈值, 调整系数后的分类准确率	47

第1章 绪论

1.1 课题背景

概念是人们进行思想活动和知识积累的基本单元。在人的思维活动中，有很多的模糊概念。比如漂亮、大小、胖瘦、年轻等。很多情况下，我们很难直接判断一个人是漂亮的或丑陋的。因为漂亮这个概念是模糊的，不同的人有不同的定义方式，而且每个人都有令人欣赏的地方也有令人讨厌的方面，因此很难用经典的集合论来解决此类问题。由于概念本身是界限不清晰、模糊的，我们不能直接指定一个对象是属于或不属于这个概念。1965 年伯克利加利福尼亚大学电机工程与计算机科学系 L.A.zadeh 教授首先提出了模糊集的概念^[1]，用模糊隶属度表示对象与概念之间的关系。在模糊集中，模糊隶属函数的产生是一个重要的问题，是模糊集理论的基础，可以用于表示信息的不确定性，同时是概念语义表示的一种形式。模糊隶属函数有多种产生方法，隶属函数的产生方法对实际应用效果产生很大的影响。但没有一些通用的准则指导我们选择合适的模糊隶属函数。在传统的模糊领域的研究中，研究人员根据问题的性质和专家经验自定义模糊隶属函数，不同的人对模糊概念的认知是不一样的，该方法具有较强的主观性，扩展性不强，适用于低维的小数据集。但却很难解决高维数据中存在的一些问题。我们希望模糊隶属函数的确定能够较为真实客观地反映数据的内在规律，尽可能减少主观因素的影响。

现在处于信息膨胀的时代，每天都会产生大量的数据，我们希望构建数据和概念之间的联系，同时体现它们之间的语义关系。一个对象和一个概念在语义空间中距离较短，他们之前的联系就比较强。人类视觉系统的研究表明^[2]，在人类感知中只有小部分的神经元发挥作用。也即一个对象只是和部分概念有联系，和大部分概念的联系较弱或无联系，这符合人类的认知行为。在支持向量机 SVM 中^[3]，只有少数被称为支持向量的样本决定分类超平面，也即少数的训练样本确定了最终的模型。有了这种联系，我们可以做分类、聚类、回归等方面的应用研

究工作。比如图像分类、文本聚类和市场趋势的预测等。

模式分类一直是机器学习中一个重要研究方向，也是最近的一个研究热点。在实际生活中，模式分类的应用场景随处可见。在大型活动、人群密度比较高的场所，利用身份识别技术鉴别危险分子，比如人脸识别、语音识别和虹膜识别等。在工作中，可以利用指纹识别进行签到，节省人力成本。在医疗诊断中，根据患者的患病历史信息帮助医生进行诊断。在网上购物时，可以根据物品的购买信息和用户的评价，利用模式分类算法筛选优质物品。在银行贷款中，可以根据客户的消费信息构建客户的信用分类模型，对客户的贷款风险大小进行分类。每天互联网中都会出现大量的新闻，可以根据新闻的内容确定新闻的类别，给用户推荐一些他们可能感兴趣的内容。我们每天都会接受大量的电子邮件，其中有部分是广告、推销等无关垃圾信息，可以利用模式分类算法对电子邮件进行分门别类。可见，模式分类算法极大地方便了我们的生活，这一直也是各大研究机构的一个研究热点，很多大型公司也在此方面投入了巨大的人力和财力成本。

模式识别的一般过程如图 1.1 所示。首先我们采集信息，比如利用相机采集图像信息，利用录音设备采集语音信息，根据用户网上行为日志整理用户数据等。通过情况下，我们不是在理想的环境中采集信息的，所得的数据含有噪声，因此需要预处理过滤数据中的噪声和规整化数据。对于语音和图像等信号，可以设计特定的滤波器过滤信号中的噪声。对于用户行为数据，利用反作弊算法检测用户作弊数据，进而过滤这部分数据。通过不同的设备采集到的信号强度有偏差，把数据化为零均值单位方差或单位向量等进行归一化。第三步是提取数据的特征。由于计算机很难直接理解一段语音或一副图片，提取表示这些信号的关键特征可以方便计算机分析处理。对于语音信号，可以提出能量、基频和共振峰等特征。对于图像数据，可以提取颜色、纹理和形状等特征。对于文本数据，我们可以用 TF-IDF 表示文档。随着提取特征的数目逐渐增加，数据表示越来越准确。但是特征之间可能是相关的，当特征增加到一定程度，反而不利于计算机的处理。在高维空间，会出现维灾难。对于高维的数据，需要利用降维方法把高维的数据变换到低维空间中，同时需要保持数据的原有的结构特性或者更方便后期算法处理。

针对不同场景的问题，研究者们提出了很多降维算法进行处理。之后，我们把数据集分为训练集和测试集，利用训练集学习一个机器学习模型。利用测试集测试模型的性能。

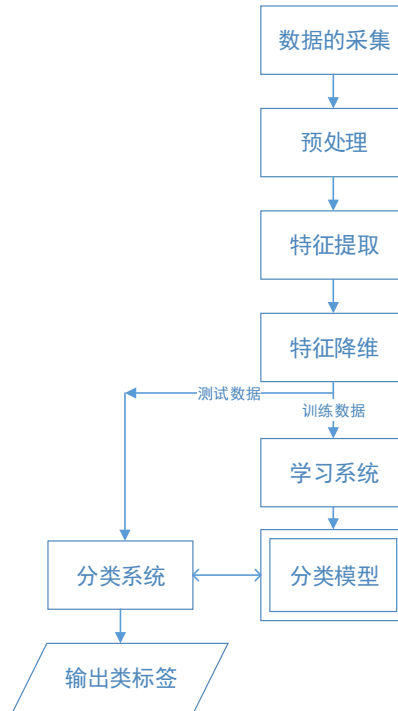


图 1.1 模式分类的一般过程

综上所述，在信息时代的背景下，针对大数据如何产生一个合适的隶属函数是一个重要的问题。同时，模式分类和我们的生活息息相关，设计一个性能较好的模式分类算法是一个重要的研究课题。在大数据的时代中，丰富的数据资源为研究提供了方便，可以有效地验证算法的性能。为此，我们提出了模糊重构算法获取模糊隶属函数，根据隶属函数进行分类应用的研究。

1.2 国内外研究现状

模糊隶属函数是模糊集理论中最为基本的一个内容。应用模糊数学解决实际问题中，确定合适的模糊隶属函数至关重要。在模糊集理论中，由于模糊隶属函数的重要作用，研究者们对其各个方面进行了深入的研究。Medasani 等总结了多种函数模糊隶属函数产生方法^[4]，包括基于直觉观察的方法、启发式的方法(把模

糊隶属函数定义为三角形、梯形、S 形等)、基于直方图转换的方法、模糊 KNN 和模糊 C 均值(Fuzzy C-Means)等方法, 其中基于直觉的隶属函数应用最为广泛, 但它不够灵活, 未利用训练数据, 不适应于高维的数据集。隶属函数的确定方法对实际应用效果产生很大的影响, 没有一个方法适用于所有的问题, 针对具体的问题选择相应的隶属函数。Ananthi 等讨论了选取不同的形状的隶属函数, 对灰度图像分割的影响^[5]。图像分割是图像分析中一项基础和必要的研究工作, 把一副图像分成多个有意义的部分, 如目标对象和背景。根据基于受限的等价函数、S 型和 Z 型、指数和 Log 等五种模糊隶属函数确定最终的直觉模糊隶属函数。由于用户在选择合适的隶属函数时候缺乏充足的知识, 因此用户会犹豫选择哪个隶属函数。根据该犹豫度进行建模, 优化求解图像分割的阈值。在图像分割之后, 图像的识别也是一个重要的问题。Hanmandlu 等进行了基于指数隶属函数的手写数字识别实验^[6]。把每个数字分成多个小区域, 该方法把每个数字图像分成 24 个区域, 提出每个区域的特征。指数隶属函数有两个参数确定, 根据熵优化求解这两个参数。该方法在印度手写数字数据集上取得较好的识别效果。不同的人书写方式不同, 以自动化或半自动化的方式识别手写数字被称为光学字符识别(OCR), 在生活中有很多的实际应用, 比如信封上手写邮编的识别。

由于模糊集可以较好地建模信息的不确定性和模糊性, 因而可以用于设计分类系统^[7]。基于模糊 If-Then 规则的模糊分类系统已得到广泛的应用^[8], 构建分类规则前件需要指明所使用隶属函数。假设使用模糊分类系统解决 m 维的空间中 c 类样本的分类问题, 每条规则可以写成如下的形式。

R_i : if x_1 is A_{i1} and x_2 is A_{i2} \cdots and x_m is A_{im} then x is the class C_i with CF_i

其中 R_i 表示第 i 条规则, x 是 m 维样本。 A_{ij} 是规则的前件, C_i 是判定该样本所属的类, $CF_i \in [0,1]$ 是规则权重。如果我们把样本空间的每个维度分成 a 段, 那么共有 a^m 规则。模糊规则数随着维度指数增加, 一般使用遗传算法进行规则的筛选。但是由于存在一些 don't care 条件, 规则前件的数量可以减少, 规则数目也可以降低。前件短和规则少是用户比较容易理解的。把最大化分类准确率和最小化规则数目作为适应性函数, 利用遗传算法选择合适的规则。隶属函数的选取

对最终结果有较大的影响^[9]。并不是每条规则的权重都一样，Ishibuchi 等提出从训练数据集学习每条规则的权重^[10]，调整规则的权重相当于改变规则的前件中隶属函数的参数，但是比较容易调整。基于区间值的模糊集能够更好地建模不确定性，基于区间值的模糊分类系统使用参数化的方法建模，在分类问题中取得了较好的实验效果^[11]。

不管在模糊领域还是在机器学习领域，模式分类都是一个重要的研究课题。KNN 是数据挖掘中常用的十大算法之一^[12]，由 Cover 和 Hart 在 1967 年首次提出^[13]。由于形式简单易于理解，在实际中应用较为广泛。KNN 需要计算测试样本相对所有训练样本的距离，当训练样本较大时，计算量较大。苏毅娟等^[14]提出利用聚类算法对训练样本分块，然后选择距离测试样本最近的块进行 KNN 分类。Keller 等^[15]把模糊集理论引入到 KNN 产生了模糊的 KNN 算法。对 KNN 进行简单的修改，利用训练样本周围样本的类标签计算它相对各类的隶属度。根据测试样本周围最近的 K 个训练样本的距离和这 K 个样本的隶属度，求出测试样本相对各类的隶属度，这样我们就可以得到模糊的 KNN。Derrac 等总结了十多种模糊 KNN 算法(FKNN)^[16]，分别在模糊隶属函数的确定、相似性的计算和决策规则等方面进行了改进，在多个数据集上进行了对比实验。Chen 等^[17]使用模糊 KNN 构建了有效的帕金森疾病诊断系统。为了提高系统的诊断准确率，需要对数据进行主成分分析 PCA 预处理，相对比基于 SVM 的方法，模糊 KNN 取得了较好的分类效果。KNN 和 FKNN 中的参数取值范围较小，鲁棒性较差。

基于人类视觉认知系统中稀疏编码机制^[18]，稀疏表示在机器学习、图像分析和视频处理等方面得到深入的研究。Wright 等^[19]首次将稀疏表示(SRC)应用人脸识别中，并取得了较好的效果。训练样本集作为字典，一个测试样本用一组字典稀疏线性表示，根据同类重构的最小残差进行分类。基于 SRC 的思想，学者们提出了很多人脸识别算法。Tang 等^[20]利用稀疏表示进行人脸分类，通过调整非零元系数的分布，使得尽可能用同类样本表示。在这些算法中，线性回归分类(LRC)^[21]相对比较简单，而容易实现。Zhang 等^[22]提出了基于联合表示的最小化方差算法(CRC_RLS)，认为在全部训练样本的联合表示对分类效果起到重要的作用，而不

是稀疏性决定分类的效果。LRC 和 CRC_RLS 都是使用的 l_2 -norm, 具有解析解, 求解速度更快。同时具有和 SRC 相当的分类准确率。由于同一类的人脸图像具有不同的姿势、光照和表情, 通过从原始人脸中产生虚拟人脸表示这种潜在的变化, 增加字典的多样性, zhang 等^[23]结合 LRC 和 CRC_RLS 进行人脸识别。

利用稀疏表示可以用于解决很多实际问题。Yang 等^[24]利用稀疏表示从原始的低分辨率的图像恢复为高分辨率的图像。Mei 等^[25]把可视化跟踪问题看作为在模板子空间中寻找近似稀疏线性表示, 其中模板包括目标模板和简单模板, 简单模板为单位向量, 仅有一个元素为 1, 其余元素为 0。Liu 等^[26]利用组稀疏方法进行交通信号识别, 限制相似的样本具有相似的稀疏编码。

字典并不一定由训练样本组成, 我们可以通过其他方式学习字典。Aharon 等^[27]提出了 K-SVD 算法, 是一般化 k-means 算法, 固定字典优化求解稀疏表示系数和固定系数更新字典, 这两步不断地迭代更新, 在更新字典时也改变相应的系数, 加快了优化速度。Ma 等^[28]基于低秩矩阵恢复学习具有判别性的字典, 从原始信号中去除噪声的同时优化字典原子重构去噪后的信号。一般的字典学习方法认为一个字典原子和一个类别相关, Yang 等^[29]提出了隐字典学习算法, 不仅需要学习一个字典, 而且需要学习它与每个类相关的权重, 同时保证同类样本具有相似的编码。

1.3 本文的主要研究内容与贡献

为了能够从数据中学习隶属函数, 本文提出模糊重构算法。在语义空间, 一个字典(训练样本)可以视为语义类标签的原型, 如果一个字典和测试样本比较相似, 那表明测试样本和该字典属于同类的可能性较大, 在重构时该字典的贡献度也较大。在字典上进行重构, 通过求解此二次优化问题获得一个模糊隶属度。模糊隶属度相当于模糊规则中的权重:

If d_i 相对测试样本 y 的重构系数为 u , then y 相对 $\text{Label}(d_i)$ 隶属度为 u 。

其中 $\text{Label}(d_i)$ 表示字典 d_i 的类标签。通过规则的合成, 我们可以获得测试样本相对各类的隶属度。同时也可以根据隶属度获得一个重构结果。根据最大隶属度原则, 对测试样本进行分类。但由于字典数目较大, 求解速度较慢。每个测试

样本必属于一个类，也即相对该类的隶属度较大，相对其他类的隶属度接近于 0。此时表明大部分字典并没有参与重构，只有部分字典参与重构，因此我们可以在重构之前选择最为相似的 K 个训练样本作为字典，这样大大提升优化速度。这就是基于近邻选择的快速模糊重构算法，可以快速地从数据中学习一个合适的模糊隶属函数。如果我们追求准确率，可以选择一个较大的 K 。如果追求速度，可以选择一个较小的 K 。而且 K 可选取范围较大。基于近邻选择的模糊重构较好地平衡了准确率与优化速度之间的矛盾。这也是一般化的 KNN 算法，当所有的重构系数都相等时退化为普通的 KNN。但是相对比 KNN，基于近邻选择的模糊重构更容易确定一个合适参数 K 。在各类标准数据集上，该算法都取得较好的分类效果。

模糊重构算法及其分类应用研究的主要贡献如下：

- (1) 从数据中学习模糊隶属函数。基于直觉方法确定模糊隶属函数，较为主观。

不同的人对同一个模糊概念有不同的认识。确定隶属函数需要客观地反映数据的本质规律，减少主观因素的影响。模糊重构算法是从数据中学习隶属函数，较为客观地刻画了数据之间关系。

- (2) 根据相对各个概念的隶属度，可以获得一个重构结果。若一个样本 y 相对各个概念 D_i ($i=1,2,\dots,n$) 的隶属度为 μ_i ，那么我们可以获得一个重构结果

$$\text{果为 } y = \sum_{i=1}^n \mu_i D_i \text{。}$$

- (3) 模糊重构可以解决分类问题。模糊重构可以获得测试样本相对各类的隶属度，我们可以直接根据最大隶属度原则进行分类。也可以根据重构系数设计一个分类器。模糊重构中字典较为庞大，但是重构系数是稀疏的，说明只有小部分字典参与重构，因此我们可以选择较为相似的部分字典作为学习的字典进行重构，字典的规模大大减小，优化的速度会提高。这也就是基于近邻选择的模糊重构，是一般化的 KNN 算法。

- (4) 对比了常见的数据降维方法的性能。数据降维是一项基础性的研究工作，在机器学习、计算机视觉、数据挖掘和数据库等领域具有重要的应用。本

文分析了主成分分析 PCA、非负矩阵分解 NMF、等测地距离映射 Isomap、局部线性嵌入 LLE 和保局部投影 LPP 五种无监督的降维方法，以及线性判别分析 LDA 有监督的降维方法。从数据可视化和分类效果两个方面评价了降维方法的性能。从中发现，模糊重构对不同的降维方法具有较强的鲁棒性。

1.4 论文组织结构

模糊重构为了解决从数据中学习模糊隶属函数的问题，该方法在分类、聚类、回归和字典学习等方面拥有较好的应用前景。本文主要关注其在分类应用中的研究，在人脸、手写数字等公开的标准数据集上都取得了较好的分类结果。本文对模糊重构在分类中的应用做了系统的研究工作，全文共分为 5 部分。每部分主要内容如下：

第一章是绪论。首先介绍了选择该课题的背景。模糊隶属函数是模糊集理论的一个基本概念，对模糊数学的具体应用有重要的影响。模式分类是模糊领域和机器学习相关领域的一个重要问题，因此模糊重构算法及其分类应用的研究具有重要的意义。然后对模糊隶属函数的产生方法和应用以及模糊重构相关的研究状况进行了详细的介绍。分析对比了各种模糊隶属函数产生方法性能，及其在具体场景中的应用，并重点介绍了模糊分类系统。综述了 KNN、模糊 KNN 和稀疏表示相关的研究工作。最后总结了本文的主要研究内容和贡献，并介绍全文的组织结构。

第二章介绍了常用的数据降维方法并分析了降维效果。本章主要介绍了主成分分析 PCA、非负矩阵分解 NMF、等测地距离映射 Isomap、局部线性嵌入 LLE 和保局部投影 LPP 五种无监督的降维方法，以及线性判别分析 LDA 有监督的降维方法。每种方法各自的优缺点，各自有针对性的应用场景。从数据可视化和分类效果两个方面评价了降维方法的性能。同时实验表明针对利用不同降维方法降维后的数据，基于近邻选择的模糊重构的分类效果始终优于 KNN。

第三章介绍了模糊重构算法及其在分类中的应用。本章开始介绍了模糊重构的两个主要相关工作，KNN 和模糊 KNN、稀疏表示 SRC。为了加快优化速度，

分析重构系数的特点，提出基于近邻选择的模糊重构，是一般化的 KNN 算法，也是一种模糊的 KNN。但是相对模糊的 KNN，该方法是根据最小化重构误差的方法确定相对各类的隶属度，K 的选取较为简单，分类效果比 KNN 较好。

第四章是实验部分。此部分是关于模糊重构在分类应用中的实验效果对比分析。针对人脸、手写数字等不同类型的标准数据集，对比了保序模糊重构、模糊重构、稀疏表示 SRC、KNN 和模糊 KNN 五种算法，分析各个算法的性能。其中模糊重构在很多情况下取得较好的分类效果。通过实验，说明基于近邻选择的模糊重构比较容易确定一个合适的 K，并讨论了确定 K 的方法。通过分析重构系数，发现了当测试样本相对各类最大的两个隶属相差较少时，容易被错误分类。如相对各类的最大隶属度为 u_1 ，次大隶属度为 u_2 ，当 $u_1/u_2 < 2$ 时，我们很难以较大的确信度判断该样本属于类 1。当两者相差较大时，容易区分。并讨论了重新调整重构系数的方法。

第五章是总结与展望。总结了全文，分析了本文的主要研究内容以及贡献。展望了模糊重构可能在字典学习、聚类 and 回归等方面的应用前景。

1.5 本章小结

本章是全文的绪论部分，分析了研究背景及其相关的研究现状。针对确定模糊隶属函数中存在的问题，提出了模糊重构算法。根据重构系数，将该算法应用到分类实验中。分析了全文的主要研究工作以及贡献。最后介绍了全文的组织结构。

第2章 数据降维方法

在大数据时代背景下,我们经常需要处理高维数据,比如图像、视频、音频、文档等。比如在人脸识别等图像的应用中,输入图像的像素为 128×128 ,经过向量量化后每个样本为 16384 维。相邻的像素点是平滑过渡的,他们之间具有一定的关联性。不同的人脸可以通过姿势、表情和光照等多个因素表示,使用那么高的维度表示是冗余的。在文本处理的应用中,每篇文档用单词表示的话,中文单词有几万个。国家在 1988 年公布的《现代汉语常用字表》选收了 2500 个常用字、1000 个次常用字,总共有 3500 字。用这 3500 个字可以表示文档中部分信息,但是很难准确无误地表达一个文档。一个 3500 维的样本也很难有效地分析。在英文中有 50 多万的单词,数目也非常庞大。如果用这些字和常用的词组表示文档,那么维度更高,更难以处理。词义相近的单词或语义相近的词组,他们之间具有一定的联系。同时占用计算机的存储空间较大,处理速度较慢,需要花费较长的时间训练模型,而且模型的泛化能力差。因此,我们需要减少特征之间关联性,用具有区分性的低维空间来表示样本。

高维数据不利于数据分析和可视化。大部分情况下,我们所提取的特征之间不是相互独立的。我们可以通过一定的变换,抽取相互独立的特征,以压缩的形式来表示样本。由此可见,数据降维对数据预处理是非常重要的。数据降维方法一般包括两种形式。一种是特征选择,通过一定的方法选择部分重要的特征表示样本,丢弃剩余的特征,如序列前向搜索(SFS)、序列后向搜索(SBS)和遗传算法(GA)等。这些特征仍然是原始的特征。另一种是特征变换,把高维数据变换到低维空间,从而将原始信息集中在这部分特征上。这些特征通常是原始特征的组合。本章主要介绍第二类方法。

2.1 维灾难

为了表示一个样本,使计算机更好地理解它。我们需要抽取样本的不同的特征,但是特征不是越多越好。因为不同的特征之间有一定的联系,引入这些冗余

特征，对我们理解该样本并无帮助。这是可能会引起过拟合的现象，因为特征维度越高，学习的模型越复杂，训练速度越慢，而且样本很稀疏。图 2.1 是一个曲线拟合的例子。我们用模拟的方法产生一个正弦曲线，并加入一定的随机噪声，

用多项式 $y = \sum_{i=0}^m w_i x^i$ 拟合正弦曲线上的点。从图 2.1 中可以看出，刚开始阶数较

低时，模型较为简单，拟合效果不好。当 $m=1$ ，也就是用一条直线拟合该曲线，效果较差。随着多项式阶数的增加，拟合的效果越来越好。当增加到一定程度之后，虽然能够完全拟合已知的训练样本点，但模型较为复杂，对测试样本的预测能力较差。此时便产生了过拟合现象。用十二阶的多项目拟合，明显发生了过拟合。

曲线拟合也可以看做一个线性回归的例子。每个样本 $x_i = \{1, x_i, x_i^2, \dots, x_i^m\}$ 用 m 个多项式特征表示。刚开始随着维度的增大，回归的效果越来越好。但是维度较高时，该回归模型较为复杂，容易产生过拟合现象，泛化能力较差。

数据稀疏是高维数据的另外一个问题。例如有 100 个样本位于一维的一个区间内，把该区间 10 等分，那么每个小区间平均有 10 个样本。当位于二维的平面内，每个维度同样 10 等分，每个小区域平均有 1 个样本。当位于三维的空间中，每个小区域平均有 0.1 个样本。当在 m 维的空间中时，每个小区域内仅有 $10^{-(m-2)}$ 个样本。随着维度的增大，每个小区域内的样本数以指数的速度减少。可见，在高维空间中，样本会非常稀疏，很难表示样本的真实分布，这对数据的统计和分析带来很大的麻烦。高维空间中距离的度量也是一个问题，任意两个样本之间的距离通常相差不大。在低维中，我们可以用欧式距离有效地表示两个点之间的距离。但在高维空间中，用欧式距离很难准确地表示两个点之间的距离。这给高维空间中的计算带来很大的麻烦。

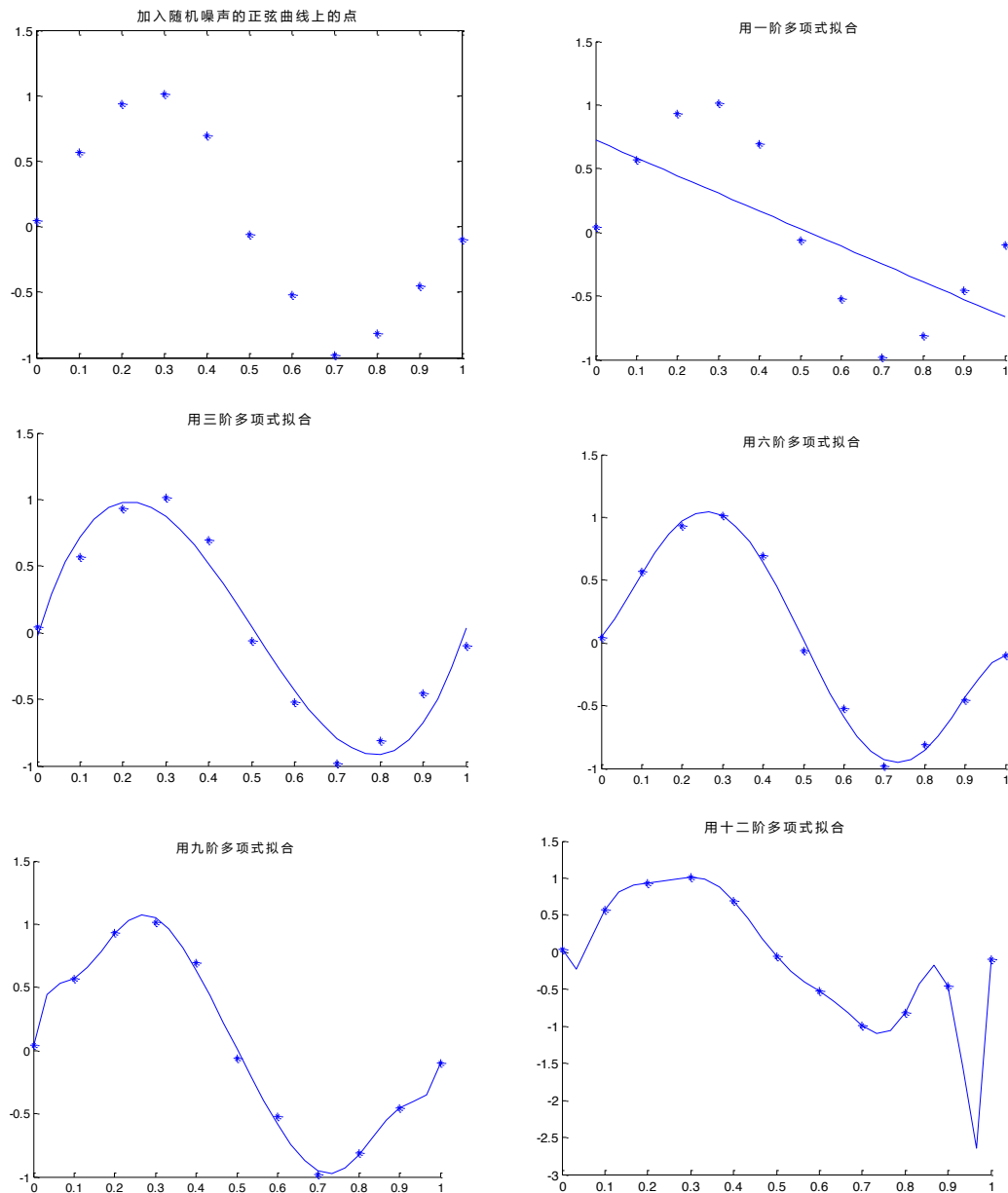


图 2.1 过拟合现象

数据稀疏和过拟合都可以通过增加样本来解决，但是在实际中，获取大量的样本需要付出很多的代价。由于过拟合学习的模型较为复杂，可以通过正则项约束模型的复杂性。也可以通过降维的方法减少特征之间的关联性，在低维空间中学习一个简单的模型。由此可见，维度并不是越高越好。

2.2 无监督的降维方法

2.2.1 主成分分析(PCA)

主成分分析(Principal Component Analysis, PCA)^[30]是一种广泛应用在数据降维、特征提取、有损数据压缩和数据可视化的技术。PCA 可以看作是寻找数据的正交投影方向,使得数据投影在低维的线性空间中的方差最大化。低维的线性空间被称为主子空间。也可以被看作最小化数据点和投影点的投影损失^[31]。图 2.2 说明了 PCA 的投影过程,数据向 u_1 方向投影,此时投影后的方差比向其他任何方向投影的方差都要大,投影后的数据点也即数据降维后的表示。

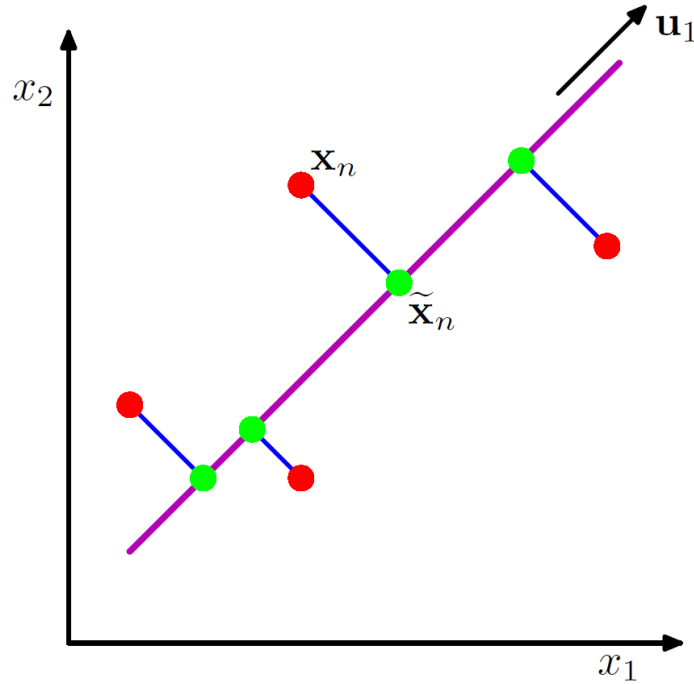


图 2.2 PCA 投影过程说明^[31]

PCA 的处理流程如下:

- (1) 归一化: 对于包含 n 个样本的数据集 $X = \{x_1, x_2, \dots, x_n\}$, 每个样本的维度为 m , 使数据集的均值变换为 0。

$$X = \frac{X - \text{mean}(X)}{\max(X) - \min(X)} \quad \text{公式(2.1)}$$

(2) 计算协方差矩阵:

$$S = \frac{1}{n} \sum_{i=1}^n x_i (x_i)^T = \frac{1}{n} X X^T \quad \text{公式(2.2)}$$

(3) 利用 SVD 求解特征值和特征向量

$$[U, T, V] = \text{svd}(S) \quad \text{公式(2.3)}$$

(4) 选择 K 个最大的特征值对应的特征向量, 进行投影。

$$Y = U(:, 1:K)^T * X \quad \text{公式(2.4)}$$

对于 $m \times m$ 的协方差矩阵, 求解特征向量的时间复杂度为 $O(m^3)$ 。从上述过程, 可以看出只需要 K 个特征向量, 并没有必要计算所有的特征向量。我们可以利用此性质加快计算速度。PCA 可以被看作为含有隐变量的极大似然模型, 用 EM 算法可以加快计算速度。在计算协方差的过程中有向量内积, 我们可以利用此特点变换成 kernel PCA^[32]。对于一些不易被线性 PCA 处理的问题, 用 kernel PCA 也许可以较好地解决。

2.2.2 非负矩阵分解(NMF)

有心理学和生理学的证据表明对物体的整体感知是基于部分感知的。非负矩阵分解(NMF)^[33]可以学习物体的部分特征, 例如人脸的部分特征嘴、鼻子和眉毛等。PCA 是学习的整体特征, 如特征脸。图 2.3 说明了他们之间的区别。这两种方法都可以认为一张人脸图像可以表示为一组基图像的线性表示。NMF 的基图像是部分脸, PCA 的基图像是特征脸, 一张正常脸的扭曲版本。

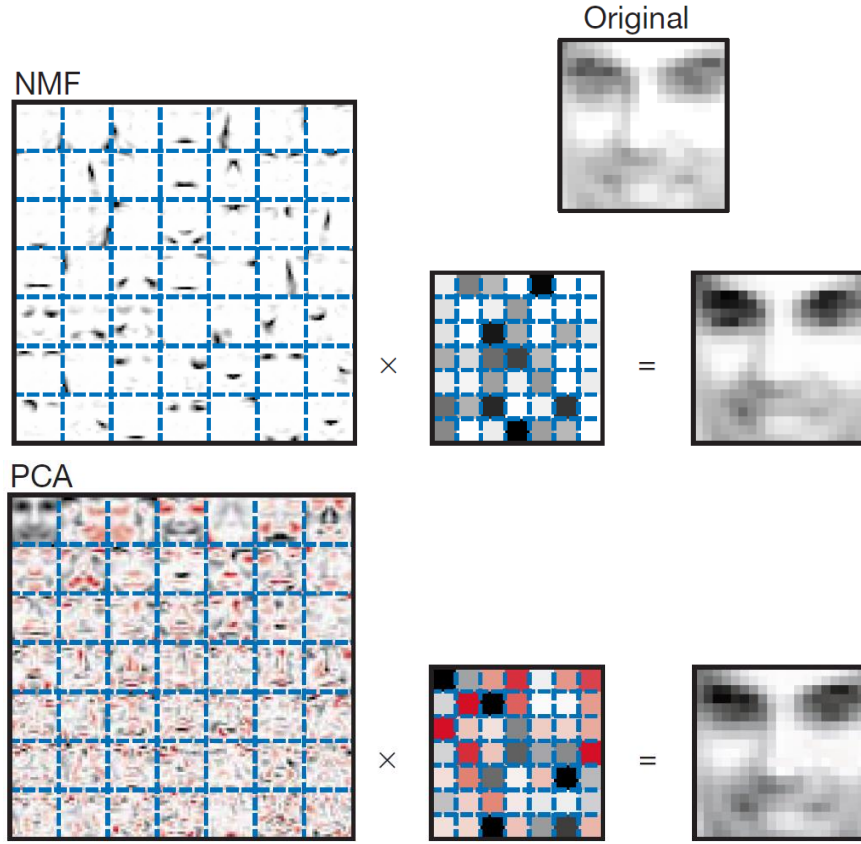


图 2.3 NMF 学习了部分脸表示，PCA 学习的是人脸的整体表示^[33]

为了分析他们之间的区别，我们可以把这两种方法统一写成公式(2.5)矩阵分解的形式。一个图像数据集表示为 $m \times n$ 的矩阵 V ，有 n 个样本，其中每一列代表具有 m 个非负特征的样本。

$$V \approx WH, \text{ 或 } V_{ij} \approx (WH)_{iu} \approx \sum_{k=1}^r W_{ik} H_{kj} \quad \text{公式(2.5)}$$

其中 W 的 r 列称为基图像， H 的每一列称为 V 的一个样本对应的编码，一张人脸图像是基图像的线性表示。

对于 PCA，要求 W 的每一列是相互正交和 H 的每一行是相互正交的。 W 的每一列也被称为特征脸，数据向此方向的投影的方差最大。 W 和 H 的符号是任意的，在这种线性表示中很多项可以相互抵消。而 NMF 要求 W 和 H 的符号都是非负的。在 NMF 中，只有不同的项相加，没有相互抵消，从而使得 NMF 学习的基图像是部分脸(基于部分的特征)。图 2.3 可以看出 W 和 H 是稀疏的。因为基图像

是部分脸特征，因此 W 是稀疏的。整个人脸是通过组合部分脸特征而获得的，但是任何人脸不必包含所有可用的部位，因而 H 也是稀疏的。

有多种方法求解 W 和 H ，Daniel 等^[34]人给出两种形式的目标函数，利用乘法和加法更新规则进行优化。其中一种目标函数定义如下

$$\begin{aligned} \arg \min_{W, H} \|V - WH\|^2 \\ \text{s.t. } W, H \geq 0 \end{aligned} \quad \text{公式(2.6)}$$

通过优化此目标函数，进而获得 W 和 H 。其中乘法的更新策略较好地平衡了优化速度和实现的复杂度。其更新策略如公式(2.7)所示。有关收敛性问题，论文中给出了详细说明。

$$H_{ij} \leftarrow H_{ij} \frac{(W^T V)_{ij}}{(W^T W H)_{ij}} \quad W_{ij} \leftarrow W_{ij} \frac{(V H^T)_{ij}}{(W H H^T)_{ij}} \quad \text{公式(2.7)}$$

2.2.3 等测地距离映射(Isomap)

在流行降维中，Isomap^[35]是一种非常有效的方法，利用局部的近邻信息学习全局的几何结构特征。PCA 是使数据投影之后的方差最大化，也就是使投影后的数据尽可能得分散。图 2.4 说明了在流行上较远的两个点，在欧式空间中距离可能较近。PCA 使用欧式距离，很难发现数据的流行结构。Isomap 是使用的测地距离，可以很好地检测这种内在的结构。对于两个近邻的点，可以用欧式距离近似表示测地距离。对于较远的两个点，可以用一系列的这些近邻点之间的最小路径和表示测地距离。当样本点的密度越高时，这种近似表示越准确。

Isomap 是保测地距离的非线性降维方式，在高维中两个点的测地距离比较大，在降维后的低维空间中他们之间的距离也要尽可能得大。Isomap 降维过程如下：

- (1) 构建邻接图，检测相邻的数据点。我们可以把一个点的固定半径

$$\|x_i - x_j\|^2 < \varepsilon \text{ 内所有的点或 } k \text{ 近邻的点定义为邻接点，他们之间的距离 } d(i, j)$$

定义为权重。

- (2) 计算测地距离，图中两个点之间的最短路径表示他们之间的测地距离。
- (3) 计算数据的低维表示，利用经典的 MDS 算法计算低维嵌入表示。

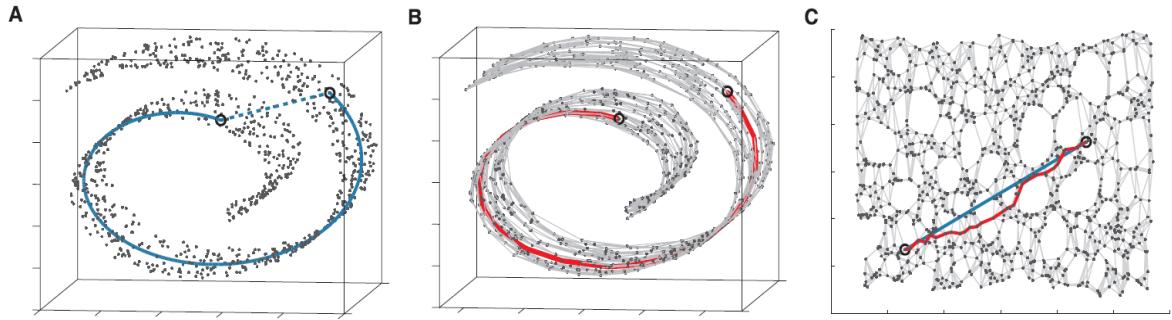


图 2.4 Isomap 的处理过程^[35]。A 说明对于非线性流行上的两个点，高维中的欧式距离无法像测地距离一样较好地反映他们之间的距离。B 说明流行上的两个点之间的测地距离可以用两点之间的最短路径较好地逼近。C 是降维后的数据，Isomap 较好地保持了近邻图中两个点之间的最短路径。

2.2.4 局部线性嵌入(LLE)

局部线性嵌入 Locally Linear Embedding (LLE)^[36]和 Isomap 是 2000 年发表在 science 上的两篇论文，都是为了解决非线性流行降维的问题。Isomap 是从全局上考虑的，需要计算任意两点之间的最短路径。而 LLE 利用局部结构恢复数据的全局非线性结构。假设在流行上有足够的样本，每个样本点和它的近邻点处在流行中的一个小区域上，每个样本可以用近邻样本进行重构。带约束的目标损失函数如下：

$$\begin{aligned} \arg \min_W \sum_{i=1}^n \left| x_i - \sum_{j=1}^n W_{ij} x_j \right|^2 \\ \text{s.t. } \sum_{j=1}^n W_{ij} = 1 \end{aligned} \quad \text{公式(2.8)}$$

W_{ij} 表示 x_j 对 x_i 样本重构的权重。若 x_j 不在 x_i 的近邻点，则 W_{ij} 为 0。也即只有近邻点参与重构。LLE 是一种保近邻的映射算法，高维数据 X 映射到低维 Y 中，仍然能够保持流行的结构。可以通过最小化下面的目标函数获得数据的低维表示。

$$\varepsilon(Y) = \sum_{i=1}^n \left| y_i - \sum_{j=1}^n W_{ij} y_j \right|^2 \quad \text{公式(2.9)}$$

通过固定 W ，优化 y 获得数据的低维表示。图 2.5 展示了 LLE 降维过程。LLE

降维的三个步骤如下：

- (1) 计算每个样本点的近邻点。可以把该样本的 K 个最近邻点或固定半径内所有的点作为近邻点。
- (2) 计算重构权重 W 。根据 X , 优化公式(2.8)获得权重 W 。
- (3) 计算数据的低维表示 Y 。固定 W , 优化公式(2.9)获得数据的低维表示 Y 。

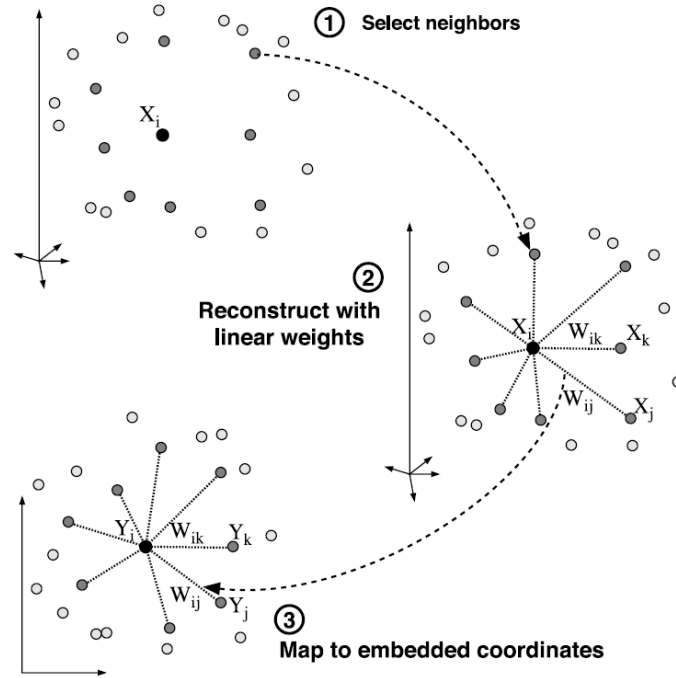


图 2.5 LLE 降维过程^[36]。①.计算每个样本点的近邻点 ②.计算重构权重 W ③. 计算数据的低维表示 Y 。

2.2.5 保局部投影(LPP)

LPP^[37]是一种线性降维方法，具有保近邻性，高维中的近邻点在低维中尽可能是近邻点。LPP 可以根据包含近邻信息的相似性矩阵确定变换矩阵，对于新的样本点，可以直接利用此变换矩阵投影到低维空间。由于保近邻性和利用近邻点构建相似性图矩阵，LPP 也可以很好地解决流行降维的问题。

LPP 的处理过程如下：

- (1) 构建邻接图 G 。

如果 x_i 和 x_j 是近邻的，他们之间就有一条边。和 Isomap 类似，也有两种

定义方式。 $\|x_i - x_j\|^2 < \varepsilon$ 或一个样本点是另外一个样本点的 K 近邻点时，他们之间就有一条边相连。

(2) 计算图的权重 W 。

W 是一个稀疏对称的矩阵，如果两个点没有边相连，他们之间的权重定义为 0。如果有边相连可以按照下面的两种方式计算。

a. Heat kernel 方法，如果 x_i 和 x_j 有边相连，他们之间的权重为

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$$

b. 简单直接的方法，如果 x_i 和 x_j 有边相连，他们之间的权重为 $W_{ij} = 1$

(3) 特征映射。

计算下式的特征值和特征向量问题。

$$XLX^T \alpha = \lambda XDX^T \alpha \quad \text{公式(2.10)}$$

其中 D 是对角矩阵， $D_{ii} = \sum_{j=1}^n W_{ij}$, $L = D - W$ 是一个 Laplacian 矩阵。

(4) 计算低维表示。

公式(2.10)的特征向量根据特征值进行排序，组成了变化矩阵 A 。对于一个样本点 x_i ，对应的低维 $y_i = A^T x_i$ 。

原文中给出了详细的推导过程，类似于 PCA 的推导过程。我们也可以把 kernel 方法利用在 LLP，产生了 kernel LPP，可以更好地解决非线性降维的问题。

2.3 有监督的降维方法

上面一节主要介绍了无监督的降维方法，按照一定的规则使高维数据投影到低维空间中，没有利用数据的标签信息。本节将会介绍一种有监督的降维方法——Fisher 线性判别分析(LDA)。

2.3.1 Fisher 线性判别分析(LDA)

前面所介绍的数据降维方法，其投影方向不一定对分类有很大的作用。例如

PCA 沿主成分方向投影使得投影后的数据尽可能散开,可能会丢失一些区别性的信息。例如在识别字母“O”和“Q”,用 PCA 方法能够发现这两个字母的相似之处,但有可能丢弃具有区别“O”和“Q”的一撇。LDA^[38]利用了数据的标签信息,用来寻找有效区分不同类的方向。图 2.6 说明了选择不同方向投影的重要性,选择主成分方向对区分不同的样本不一定是最有利的。

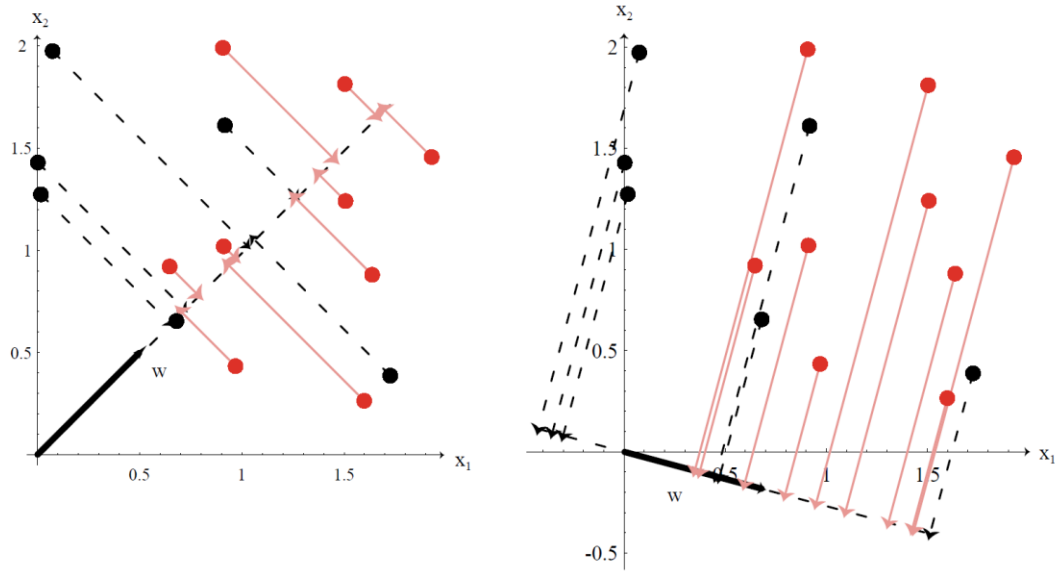


图 2.6 PCA 和 LDA 投影效果对比^[38]。同一个数据集向不同的方向投影,左图利用 PCA 投影,右图利用 LDA 投影,右图投影后的点更容易分开。

首先我们考虑较为简单的二维情况,有 n_1 个 C_1 类的样本,有 n_2 个 C_2 类的样本。对于 m 维的样本,利用下式投影到一维中。

$$y = w^T x \quad \text{公式(2.10)}$$

样本均值之差可以用来衡量两类样本的分析程度。两类样本的均值定义如下:

$$m_1 = \frac{1}{n_1} \sum_{x \in C_1} x, \quad m_2 = \frac{1}{n_2} \sum_{x \in C_2} x \quad \text{公式(2.11)}$$

投影后的样本均值为:

$$\overline{m}_1 = \frac{1}{n_1} \sum_{y \in C_1} y = w^T m_1, \quad \overline{m}_2 = \frac{1}{n_2} \sum_{y \in C_2} y = w^T m_2 \quad \text{公式(2.12)}$$

为了使投影之后的样本点具有较好的区分性,需要确保同样样本的聚集程度

比较大, 不同类样本之间尽可能分开。因此, 可以定义如下的目标函数:

$$J(w) = \frac{(\overline{m1} - \overline{m2})^2}{s_1^2 + s_2^2} \quad \text{公式(2.13)}$$

其中 s_1 和 s_2 分别为类 C1 和类 C2 中同类数据的变化(方差的 n 倍)。

$$s_1^2 = \sum_{y \in C1} (y - \overline{m1})^2, \quad s_2^2 = \sum_{y \in C2} (y - \overline{m2})^2 \quad \text{公式(2.14)}$$

把上面各式带入到公式(2.13)整理可得

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \quad \text{公式(2.15)}$$

其中 S_B 是不同类的之间变化矩阵, S_w 是同类样本之间的变化矩阵之和(协方差矩阵的 $n-1$ 倍)。

$$S_B = (\overline{m2} - \overline{m1})(\overline{m2} - \overline{m1})^T \quad \text{公式(2.16)}$$

$$S_w = \sum_{x \in C1} (x - m1)(x - m1)^T + \sum_{x \in C2} (x - m2)(x - m2)^T \quad \text{公式(2.17)}$$

通过优化求解公式(2.15), 可以获得

$$w = S_w^{-1}(m1 - m2) \quad \text{公式(2.18)}$$

对于 $K(K>2)$ 类的数据降维问题, 我们需要计算 K' ($K'<K$) 个 w 向量。这些 w 向量构成变换矩阵 W , 样本 x 降维后的表示为

$$y = W^T x \quad \text{公式(2.19)}$$

2.4 降维效果的评估

根据问题的性质, 确定评价降维性能的方法。在数据可视化中, 我们希望去除无关属性的影响, 在低维空间中能够更好地观察数据的分布规律。在机器学习领域, 不仅要求去除一些无关属性, 而且希望能够保持数据的内在结构规律, 在低维中能够更好地进行分类、聚类或回归等。本节从可视化效果和分类效果两个方面评价前两节所描述的降维方法。

2.4.1 可视化效果

高维的数据不利于可视化，可以将数据投影到二维或三维空间中，观察数据的分布。本节选择 Extended Yale B 数据集中的两个人的人脸数据进行可视化。在第四章将会详细介绍该数据集。原始样本的维度为 192×168 ，由于高维的样本计算量比较大和相邻的像素点之间有联系，可以下采样至 32×32 ，减少计算量。为了更好地观察数据点的分布，仅可视化两类人脸数据。

利用前面介绍的六种降维方法(包括主成分分析 PCA, LLE, Isomap, LPP, NMF 和线性判别分析 LDA)降至二维空间。图 2.7 说明了利用不同的无监督降维方法进行数据降维，在二维空间中样本的分布情况，红色空心圆表示一类数据，蓝色实心圆表示另外一类数据。从直观可视化的角度考虑，降维后的不同类的样本点距离越远，该降维方法的效果就越好。在 LLE 中，利用 $K(K=5)$ 近邻进行重构。在 Isomap 中，也是根据 $K(K=5)$ 确定邻接点。在 LPP 中，根据 $K(K=3)$ 近邻确定邻接图，利用热核函数($t=2.0$)确定边的权重。从图 2.7 可以看出，利用 PCA 降维之后的数据方差比较大，没有流行降维方法(LPP, LLE, Isomap)的效果好，不同类的数据耦合在一起了。这些流行降维方法可以保持数据流行的原始结构，使得降维后的不同类的数据点比较分散。其中非线性 LLE 和 Isomap 的降维效果稍微优于线性的 LPP。LPP 产生一个投影矩阵，新的样本可以根据该投影矩阵进行降维，计算量较小。但是对于新的样本，LLE 和 Isomap 不能利用已有的计算结果，需要和原始数据一起重新计算，计算量较大。

图 2.8 展示有监督的降维方法(LDA)的降维效果。对于包含 C 类的样本，最多只能向 $C-1$ 维的空间做投影。左图显示了包含两类样本的数据集投影到一维空间的情况，从中可以看出两类数据完全可以分开，而且中间有很大的间距。为了显示这些样本投影到二维空间的情况，在原始的数据集中另外加入一类数据，右图显示了这两类样本投影到二维空间中的分布情况。从中可以发现两类样本完全分离且相距较远。对比图 2.7 和图 2.8 可以发现，引入标签的有监督的数据降维方法(LDA)比无监督的降维方法的效果好一些。

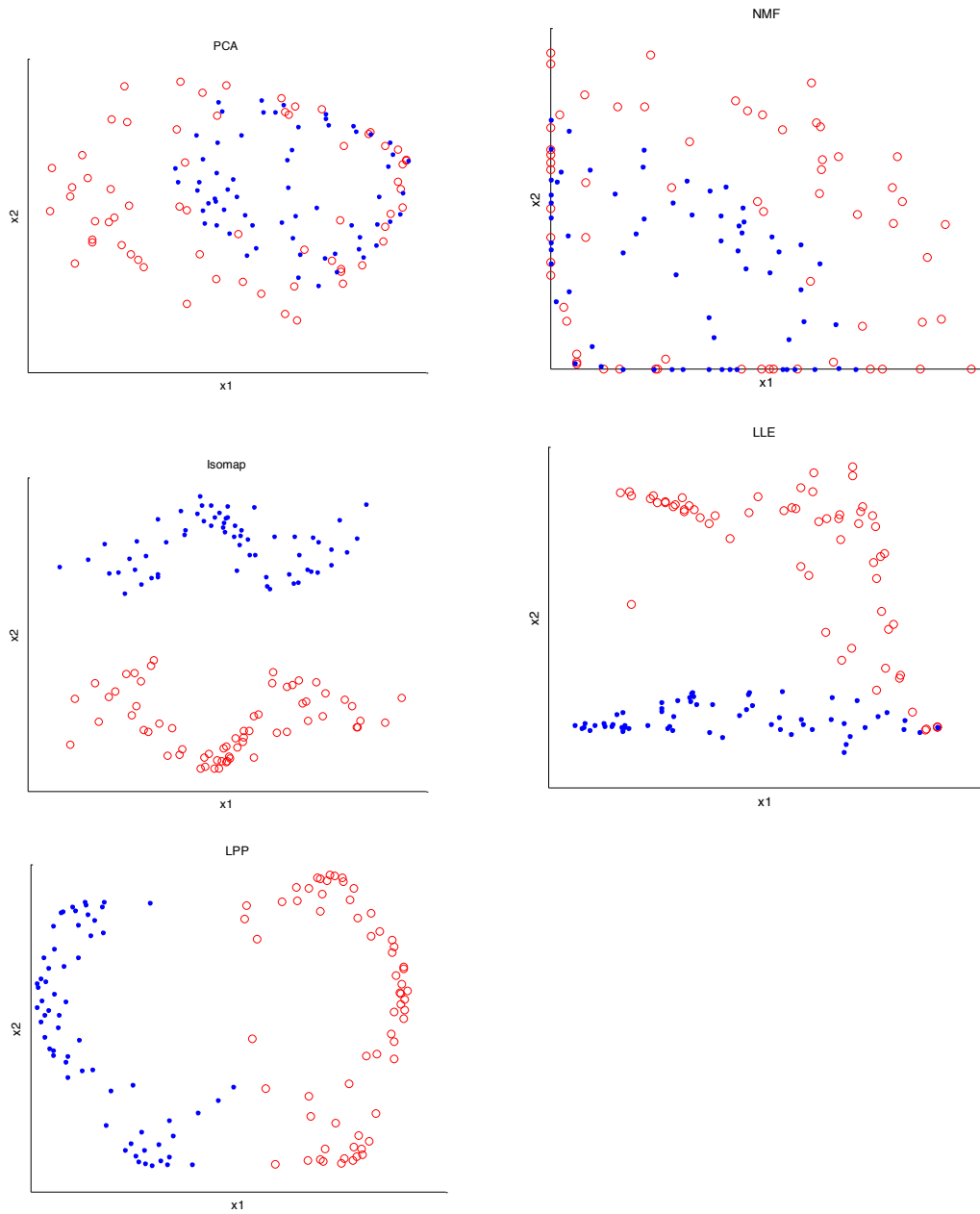


图 2.7 利用有监督的降维方法投影到二维空间中样本点的分布情况

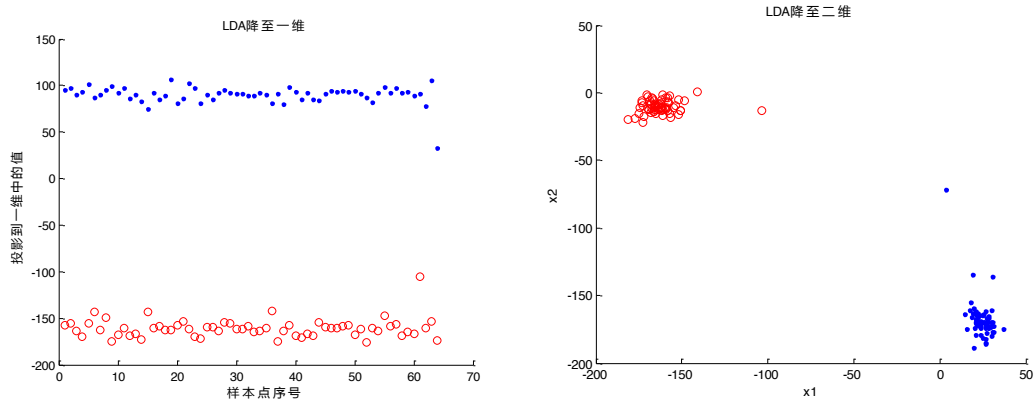


图 2.8 利用有监督的降维方法投影到二维空间中样本点的分布情况

2.4.2 分类效果

高维中存在维灾难，利用数据降维方法降到低维空间，进行分类实验。此次分类实验和上一节可视化实验使用的是同一数据集。其中一半的数据集用于训练，剩余的样本用于测试。在 PCA、NMF、LPP 和 LDA 中，利用训练样本学习变换矩阵，测试样本利用变换矩阵变换到低维空间中。LLE 和 Isomap 无法处理新的样本点，需要把训练样本和测试样本放在一起进行降维。经过零均值法(首先将每一个维度归一化至 $[0,1]$ ，然后均值化为 0)或单位向量法(所有的样本点化为单位向量)预处理之后，利用不同的降维方法把原始的 1024 维数据分别降至 16、32、64、128 维，然后使用 KNN 进行分类。其中 LDA 最多只能降至 37 维。

由于 NMF 算法中，随机初始化了 W 和 H 矩阵，有的 W 和 H 可能使得该算法收敛到局部极小值，因此我们进行了 10 次实验取分类准确率的平均值。PCA 要求投影后的数据尽可能分散，即使所有的样本归一化为单位向量，也需要把每个维度的均值化为 0。表 2.1 是数据降维后 KNN 的分类效果对比，其中 K 分别取 1、3、5、7，表中记录了最高的分类准确率。从中可以看出对样本进行单位向量的预处理比零均值法的效果好一些。有监督的方法降维方法优于无监督的降维方法。在 32 维的空间中，LDA 分类准确率相对 Isomap 提高了 21.21%，相对 PCA 提高了 115.42%，其中 Isomap 是在 32 维中分类最好的一种无监督的降维方法。有监督的降维方法考虑了样本的标签信息，使得降维后的数据更易分类。非线性的流行降维方法(如 LLE, Isomap)的分类效果优于线性的降维方法(如 PCA 和

LPP)。在 32 维的空间中，Isomap 的分类准确率相对 LPP 提高了 79.05%，相对 PCA 提高了 77.73%。非线性的流行降维方法考虑了流行的结构，测试样本和训练样本一起降维，相对线性降维方法，训练样本数目多，更好地反映了样本的分布规律，因而分类效果好。但是计算量较大。Isomap 和 LLE 都是非线性的流行降维方法，但是 Isomap 的分类效果优于 LLE，由于 Isomap 利用测地距离表示样本之间的结构信息，LLE 是通过局部的欧式距离表示全局的结构信息。一般情况，非负矩阵分解的效果好于其他的无监督降维方法，但是非负矩阵分解需要多次初始化 W 和 H ，迭代优化求解，求解速度较慢，且可能收敛到局部极小值，最终解和初始化的 W 和 H 有关。

表 2.1 数据降维后，KNN 分类效果对比

		16	32	64	128
PCA	零均值	0.3052	0.4467	0.5476	0.6063
	单位向量	0.2597	0.4500	0.6146	0.7113
NMF	零均值	0.5241	0.7550	0.8399	0.8908
	单位向量	0.5204	0.7744	0.8690	0.9107
LPP	零均值	0.5955	0.6518	0.6766	0.7022
	单位向量	0.7477	0.7891	0.8040	0.8205
LLE	零均值	0.5699	0.6261	0.6476	0.6882
	单位向量	0.7130	0.7767	0.7907	0.7932
Isomap	零均值	0.5236	0.5525	0.5616	0.5707
	单位向量	0.7767	0.7998	0.8189	0.8280
LDA	零均值	0.8850	0.9338	0.9438 (dim=37)	
	单位向量	0.9446	0.9711	0.9752 (dim=37)	

表 2.2 是利用基于近邻选择的模糊重构的分类效果对比，其中加粗的字体表示有监督和无监督的数据降维后最高的分类准确率。本次实验中所有降维方法的参数和上面实验中的一样。基于近邻选择(KNN)的模糊重构可以加快模糊重构的速度。由于表 2.2 需要进行大量的计算，我们选择一个较少的 K(K=38,样本的类别数)。从表 2.2 可以看出，线性的流行降维方法 LPP 和非线性的 LLE 和 Isomap 的分类效果基本相当，他们三者和线性降维方法 PCA 也相差无几。非负矩阵分解 NMF 的效果在无监督的数据降维方法中效果较好。和表 2.1 同样表明，有监督的降维方法优于无监督的降维方法。对比表 2.1 和表 2.2，基于近邻选择的模重构的分类准确率通常高于 KNN，它相对 KNN 对预处理方式和降维方法具有较强的鲁棒性。

表 2.2 数据降维后，基于近邻选择的模糊重构分类效果对比

		16	32	64	128
PCA	零均值	0.6079	0.7502	0.8395	0.8685
	单位向量	0.5715	0.7469	0.8321	0.8776
NMF	零均值	0.7253	0.8336	0.9166	0.9462
	单位向量	0.6937	0.8544	0.9261	0.9445
LPP	零均值	0.6352	0.6989	0.7270	0.7568
	单位向量	0.7717	0.8040	0.8255	0.8453
LLE	零均值	0.6336	0.6998	0.7188	0.7419
	单位向量	0.7750	0.8288	0.8462	0.8652
Isomap	零均值	0.6013	0.6270	0.6352	0.6534
	单位向量	0.8073	0.8420	0.8577	0.8701
LDA	零均值	0.8801	0.9636	0.9628 (dim=37)	
	单位向量	0.9289	0.9636	0.9669 (dim=37)	

2.5 本章小结

由于高维数据存在维灾难问题，本章主要介绍了常用的数据降维方法，包括主成分分析 PCA、非负矩阵分解 NMF、等测地距离投影 Isomap、局部线性嵌入 LLE 和保局部投影 LPP 五种无监督的降维方法，以及线性判别分析 LDA 有监督的降维方法。通过数据可视化和分类效果两个方面来评价数据降维方法的性能。本章使用的是人脸数据集，一般人脸位于流行上，所以流行降维方法在该数据集上取得较好的效果。针对不同的问题，不同的降维方法各有千秋，因而要针对不同问题选择相应的降维方法。基于近邻选择的模重构的分类准确率通常高于 KNN，它相对 KNN 对预处理和降维方法具有较强的鲁棒性。

第3章 模糊重构算法

本章主要介绍利用模糊重构算法求解模糊隶属函数，根据隶属函数进行分类实验。在重构的过程中，希望训练样本和测试样本越相似，它对测试样本的重构贡献度越大。为了加快求解速度，通过分析重构系数，我们提出了基于近邻选择的模糊重构。

在模糊重构以及相关的算法中，都会使用一些符号。为了方便阅读，下面对这些符号进行统一说明。假定有 $D \in R^{m \times n}$ ，其中 D 称为字典，本文是由训练样本组成， $D = \{D_1, D_2, \dots, D_c\}$ ，其中 $D_i \in R^{m \times n_i}$ ， $D_i = \{D_{i1}, D_{i2}, \dots, D_{in_i}\}$ 为第 i 类的训练样本，每类有 n_i 个样本，共有 c 类， n 个样本，每个样本是 m 维的列向量。 $y \in R^{m \times 1}$ 为测试样本， $\mu \in R^{n \times 1}$ 为重构系数。

3.1 K 近邻算法和模糊 K 近邻算法

3.1.1 K 近邻算法(KNN)

KNN 是一种基本的分类算法。由于形式简单和易于实现，KNN 在分类研究中经常作为一个对比方法。在分类时，对于给定的测试样本，根据距离其最近的 K 个训练样本的类标签，通过多数表决等方式预测测试样本的类别。

K 值的选择、距离的度量方法和决策方法是 KNN 算法的三个基本要素。 K 值较小，测试样本对近邻点较为敏感，增大了噪声的干扰，特征空间被分割的区域数较多，因此模型较为复杂，容易过拟合。 K 值较大，用较大领域内的训练样本进行预测，特征空间被划分的区域数少，因而模型简单，容易欠拟合。当 K 为训练样本数，无论输入什么样的测试样本，都将它预测为训练样本中最多的类别。针对不同的应用场景， K 值也是不一样的。在实际应用中， K 一般取较少的奇数值，可以通过交叉验证等方式确定合适的 K 。一般用欧氏距离、曼哈顿距离、马氏距离和余弦相似性作为距离的度量方式。在高维空间中，通常使用改进后的余

弦相似性度量两个样本之间的距离。KNN 的决策方式通常是多数表决和加权投票的方式。

3.1.2 模糊 K 近邻算法(FKNN)

把模糊集理论的思想引入到 KNN 中得到了模糊 KNN(FKNN)。FKNN 通过引入模糊隶属度、改进相似性计算方法、新的决策规则等方式提升 KNN 的分类效果。FKNN 不是赋予测试样本一个具体的类标签，而是指定相对每个类的隶属度。FKNN 通过加权投票的方式进行预测。 $u_i(x)$ 表示测试样本 x 相对 i 类的隶属度，与 K 个最近邻样本的距离和这 K 个样本相对每类的隶属度有关。 $u_i(x)$ 计算方式见公式(3.1)。

$$u_i(x) = \sum_{j=1}^K u_{ij} (1/d_j^2)^{1/(m-1)} / \sum_{j=1}^K (1/d_j^2)^{1/(m-1)} \quad \text{公式(3.1)}$$

其中 d_j 是测试样本 x 和第 j 个训练样本之间的距离， u_{ij} 表示第 j 个的训练样本相对类 i 的隶属度。 u_{ij} 可以通过下面的方式计算。

$$u_{ij} = u_i(x_j) = \begin{cases} 0.51 + (n_i/K) \times 0.49 & \text{当 } L(j) = i \\ (n_i/K) \times 0.49 & \text{当 } L(j) \neq i \end{cases} \quad \text{公式(3.2)}$$

其中 n_i 表示 x_j 的 K 近邻样本属于 i 类样本的数目， $L(j)$ 表示第 j 个样本的类标签。

3.2 稀疏表示 SRC

Wright 等^[19]提出基于稀疏表示(SRC)的人脸识别，一个样本可以用一组字典线性表示，这里的字典是由训练样本组成，希望表示的系数是稀疏的。根据同类样本的线性表示和测试样本之间的最小残差作为分类标准。由于字典过完备的假设，字典元素较多，测试样本用字典线性表示时，可能有无穷多个最优解。但是我们希望用尽可能少的字典表示它。如果每一类都用足够的字典样本，希望这些参与表示的字典和测试样本属于同一类。

直观上，我们通过 ℓ_0 零范数获得稀疏最优解，其中零范数是通过计算非零元的个数。最小化系数的零范数是 NP-hard 问题，虽然有一些次优解，但是效果不

是很理想。当系数是足够稀疏时， ℓ_1 范数的效果等价于 ℓ_0 零范数^[39]。

SRC 算法的处理过程如下：

算法 1：基于稀疏表示的分类算法(SRC)

(1) 按列归一化字典 D 和测试样本 y

(2) 求解 ℓ_1 最小化问题：

$$\mu = \arg \min_{\mu} \|D\mu - y\|_2^2 + \lambda \|\mu\|_1$$

(3) 计算残差 $e_i(y) = \|D\mu_i - y\|_2^2$, 其中 μ_i 表示与第 i 类相关的重构系数。

(4) 测试样本 y 的类标签： $identify(y) = \arg \min_i (e_i(y))$

当 y 属于第 i 类时，通常可以用 $y \approx D\mu_i$ 近似表示，也即表明 $\mu_j, j \neq i$ 的大部分值为 0，只有第 i 类字典对 y 的线性表示有贡献。上面带 ℓ_1 罚项的目标函数最小化问题，可以通过内点法求解^[40]，但优化速度较慢。对于规模较大的数据集，可以使用 SPAMS 进行快速求解^[41]。由于字典完备的要求，字典的规模较大，优化速度仍然较慢。

3.3 模糊重构

模糊重构可以从数据中学习模糊隶属函数。字典中的原子(训练样本)和测试样本越相似，我们希望它对测试样本的重构的贡献度越大。对于一个新的测试样本 $y \in R^{m \times 1}$ ，我们需要知道该样本相对字典 D 中每类样本的隶属度，且模糊隶属度 $\mu \in [0, 1]$ ，为此可以在字典 D 上进行模糊重构。因此，我们定义如下的带约束的目标函数：

$$\begin{aligned} \mu &= \arg \min_{\mu} \|D\mu - y\|_2^2 \\ s.t. \quad &0 \leq \mu \leq 1 \end{aligned} \quad \text{公式(3.3)}$$

通过求解此二次优化问题，最终可以获得模糊隶属度 $\mu \in R^{n \times 1}$ 。每个训练样本可以看作是一条模糊规则，重构系数相当规则的权重，该算法具有 n 个模糊规则的推理系统。但是相对模糊推理系统的规则，具有较强的可解释性，而且规则数

较少。为了获得测试样本 y 相对 c 个类的隶属度，可以通过下面的求和边界算子获得。

$$\mu_i = \min \left(\left\{ \sum_i^{n_i} \mu_{in_i} \right\}, 1 \right) \quad \text{公式(3.4)}$$

我们也可以其他算子求解同类样本内隶属度的并。不同的算子对该实验结果影响并不是特别大。根据测试样本 y 对于 c 个样本的隶属度，可以把 y 归到隶属度最大的类中。模糊重构算法归纳如下：

算法 2：模糊重构(FR)

(1) 按列归一化字典 D 和测试样本 y

(2) 求解带约束的目标函数

$$\begin{aligned} \mu &= \arg \min_{\mu} \|D\mu - y\|_2^2 \\ \text{s.t. } 0 &\leq \mu \leq 1 \end{aligned}$$

(3) 相对 c 个类的隶属度 $\mu_i = \min \left(\left\{ \sum_i^{n_i} \mu_{in_i} \right\}, 1 \right)$

(4) 把 y 归到 μ_i 最大的类中

图 3.1 展示了该算法了整体过程，左边的是测试样本，中间是由训练数据组成的字典，在中间的字典上进行重构而获得右图的重构系数。通过求和边界算子，我们可以获得如图 3.2 的隶属度。从中可以看出，该测试测样本相对所在类(类标签 16)的隶属度较大，接近于 1，而相对其他类的隶属度都较小。

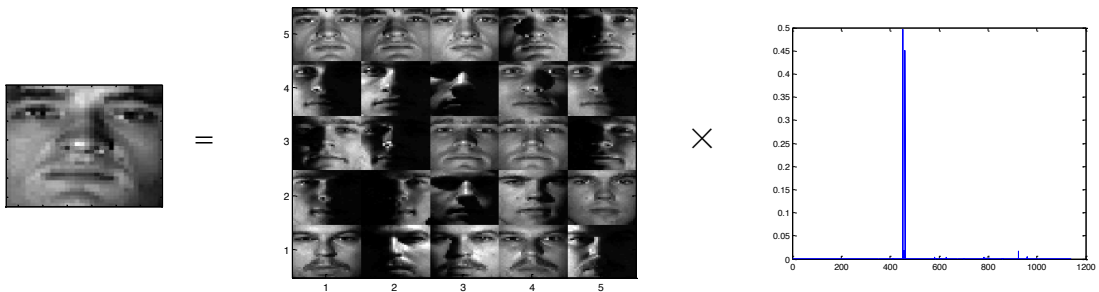


图 3.1 模糊重构算法的整体过程。左图是测试样本，中间是部分训练集，右边是重构系数。

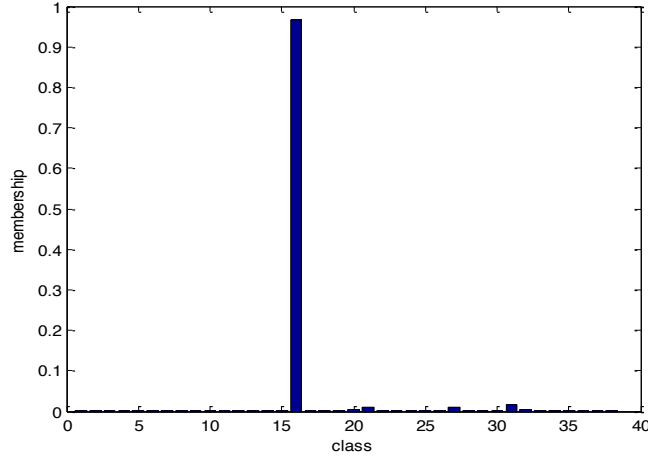


图 3.2 测试样本相对 $K(K=38)$ 类的隶属度。系数较大的样本所在的类就是该测试样本所在的类，模糊重构的方法可以正确区分该样本所在的类。该数据是 Extended Yale Face Dataset B，共有 38 类。

如果测试样本和训练样本较为相似，那么这两个样本有可能属于同一类，此时重构系数应该比较大，隶属度较高。同样的，如果两个样本根本不相似的话，那么它们不属于同一类，重构系数接近于 0。从图 3.1 最右侧重构系数图和图 3.2 中可以看出，只有部分重构系数不为 0，这些字典原子和测试样本属于同一类。很多重构系数近似为 0，这些字典原子和测试样本不属于同一类。这说明只有部分字典在重构时起作用，而且是相似度是比较高的。因此可以只选择这部分字典参与重构。有很多的方法计算相似性，我们选择两个样本之间的欧式距离度量他们之间的相似性，该方法较为简单和计算速度较快。另一方面，由于模糊重构是带约束的优化求解问题，故无解析解。利用 CVX 等优化求解器解决该问题时，优化速度较慢。

3.4 保序模糊重构

同样基于重构的思想，Tang 等^[42]提出了保序模糊重构 OPFR，用于度量测试样本与语义标签原型之间的相似性，在回归实验中取得较好的效果。“物以类聚”，相似的样本在样本空间中的距离较近，在样本重构中它的贡献度也较大。假定 D 为字典， $Label \in 1 \times n$, $Label_i$ 为实数，对应第 i 个字典的输出， $dist_i$ 为测试样本 y 与第 i 个字典之间的距离，不失一般性，对于任意的 $i(1 < i < n-1)$ ，都有

$dist_i \geq dist_{i+1}$ 。在字典 D 上进行如下的重构：

$$\begin{aligned} \mu &= \arg \min_{\mu} \|D\mu - y\|_2^2 \\ s.t. \quad \forall i, 0 \leq \mu_i \leq \mu_{i+1} \leq 1 \end{aligned} \quad \text{公式(3.5)}$$

为了方便求解此目标函数，可以假定 $v_1 = \mu_1, v_i = \mu_i - \mu_{i-1} (2 \leq i \leq n)$ ，以及 $Z = DS$ ，其中

$$S = \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ \vdots & \vdots & \ddots & & \\ 1 & 1 & \cdots & 1 & \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \in n \times n \text{ 下三角矩阵。}$$

可以把上面的优化问题转化成下面的形式。

$$\begin{aligned} v &= \arg \min_{\mu} \|Zv - y\|_2^2 \\ s.t. \quad \forall i, 0 \leq v_i \leq 1 \end{aligned} \quad \text{公式(3.6)}$$

可以像模糊重构一样利用二次优化求解目标函数，可以获得重构系数 v ，进而求得原始目标函数的重构系数 $\mu = Sv$ 。每一个字典相当于一条模糊规则，利用 $Label * \mu$ 获得回归值。保序模糊重构的处理过程可以简述成如下的形式：

算法 3：保序模糊重构(OPFR)

- (1) 按列归一化字典 D 和测试样本 y
- (2) 求解带约束的目标函数

$$v = \arg \min_{\mu} \|Zv - y\|_2^2 \quad s.t. \quad \forall i, 0 \leq v_i \leq 1$$

- (3) 获取原始目标函数的重构系数 $\mu = Sv$
- (4) 测试样本 y 的回归值为 $Label * \mu$

此算法也可以用于分类问题，只需要修改最后的决策函数。为了说明不同算法分类结果的可比性，SRC 和保序模糊重构 OPFR 利用最大同类正系数和作为分类准则。

3.5 基于近邻选择的模糊重构

因为非负限制，模糊重构使得重构系数是稀疏的，也即大部分的重构系数为 0，只有部分样本参与重构。通常情况下同类样本的相似性较大，我们希望参与重构的样本和测试样本属于同一类。图 3.2 表明，大多数参与重构的样本和测试样本属于同一类。为了利用稀疏性和解决优化速度问题，我们提出了基于近邻选择的模糊重构。首先，选择和测试样本最为相似的 K 个样本，然后在此基础上进行上述的重构。可以利用欧式距离、余弦距离、测地距离等衡量两个样本之间的相似度，本文采用了欧式距离。

算法 4：基于近邻选择的模糊重构(KFR)

- (1) 按列归一化字典 D 和测试样本 y
- (2) 选择和 y 最为相似的 K 个字典 \bar{D} 作为训练字典
- (3) 求解带约束的目标函数

$$\begin{aligned} \mu &= \arg \min_{\mu} \|\bar{D}\mu - y\|_2^2 \\ s.t. \quad &0 \leq \mu \leq 1 \end{aligned}$$

$$(4) \text{ 相对 } c \text{ 个类的隶属度 } \mu_i = \min \left(\left\{ \sum_i^{n_i} \mu_{in_i} \right\}, 1 \right)$$

- (5) 把 y 归到 μ_i 最大的类中

经过近邻选择预处理，字典的规模较小了，可以大大加快优化速度。但同时精度也会有所下降。我们只要选择一个合适 K 值，精度下降的幅度在可接受范围内。基于近邻选择的模糊重构可以看作是一般化的 KNN 算法，当重构向量系数 μ 中各元素的值都相等时，该算法就退变为 KNN 算法。当选择最近邻的模糊重构时，该算法就变成了最近邻 NN。基于近邻选择的模糊重构相当于修改了 KNN 的决策方式，利用重构的贡献度作为权重预测测试样本的类别。该方法相对 KNN 具有更好的灵活性和更高的准确率。图 3.3 表示包含两类样本点的数据分布情况，小圆圈表示第一类样本，方框表示第二类样本。从中可以说明该方法相对于 KNN 的优点。颜色的深度表示模糊重构值(隶属度)的大小。假设一个测试样本是二维

空间中无颜色的方框，它是属于第二类。如果只是用 KNN 的方法，当 $K > 4$ 时，该测试样本容易被错误分在类 1 中。基于近邻选择的模糊重构由于引入了重构思想，而且要求重构系数是非负的。该测试样本可能只需要同类的样本进行重构，图中 2 个深蓝色方框点对它的重构贡献度较大，其他点的贡献度较低。基于近邻选择的模糊重构和模糊 KNN 都能求得一个样本的隶属函数，而不是仅仅判定属于具体的某个类。但基于近邻选择的模糊重构引入了重构思想，且要求系数是非负的，并非所有的样本都参与测试样本的重构。在模糊 KNN，所有近邻样本都影响预测的效果。由此也可以看出，基于近邻选择的模糊重构拥有模糊重构的较好的分类效果，也具有 KNN 快速处理的能力。基于近邻选择的模糊重构较好地平衡了分类效率和优化速度之间的矛盾。当追求较高的识别率时，可以选择较大邻域(较大的 K)。当追求求解速度时，可以适当地选择较小的邻域(较小的 K)。

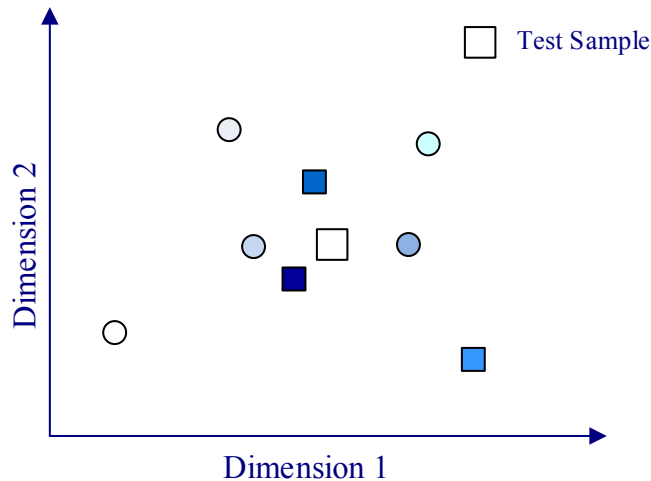


图 3.3 KFR 和 KNN 的对比。5 个小圆圈表示类 1 中样本，3 个小方框表示类 2 中的样本，中间无颜色的点表示测试样本。

在稀疏表示 SRC 和保序模糊重构中，都可以先选择和测试样本较为相似的 K 个样本作为字典，然后再此字典上进行重构。

3.6 本章小结

本章主要介绍了模糊重构算法，从数据中学习模糊隶属函数。限制重构系数是非负，只能通过字典相加，不能相互抵消，这导致了样本的稀疏表示。通过重

构系数获得模糊隶属度，根据最大隶属度原则进行分类。通过分析重构系数，我们发现参与重构的训练样本大部分来自同一类，这些样本和测试样本较为相似。同时由于字典数目较大，带约束的二次目标函数优化速度较慢。因而我们提出了基于近邻选择的模糊重构，它是一般化的 KNN 算法。保序模糊重构希望相近的样本具有较大的重构贡献度。

KNN 是常用的十大数据挖掘算法，本章介绍了 KNN 中 K 值的选择、距离的度量、决策规则的确定的 3 个基本要素。KNN 中引入模糊集的思想，便产生了模糊 KNN 算法。SRC 算法是一个通用的分类算法，测试样本通过字典线性表示，同时限制表示系数是稀疏的。

第4章 模糊重构算法的分类实验

前面一节详细地介绍了模糊重构算法。为了解决优化速度慢的问题，我们提出了基于近邻选择的模糊重构算法。传统意义上的模糊隶属函数求解方法不适用于高维数据和大数据集，但现在处于大数据时代，模糊重构方法扩展了模糊隶属函数求解方法，使得一些优化算法可以应用在模糊系统中，建立了模糊系统与大数据之间的桥梁。本章我们应用该算法解决分类问题。由于数据的维度较高，在第二章已经对比分析了不同的降维方法对分类效果的影响，模糊重构对不同的降维方法具有较强的鲁棒性，PCA 是线性的降维方法，使用较为广泛，相对非线性的流行降维方法速度较快，故本章选用 PCA 作为降维方法。在人脸、手写数字等标准的数据集上，模糊重构(FR)、基于近邻选择的模糊重构(KFR)的分类效果始终优于 KNN、模糊 KNN(FKNN)。

SRC 在分类问题中取得较好的实验效果，OPFR 在回归试验中取得较好结果。在前文中我们已经提出，OPFR 也可以解决分类问题。但对于高维大数据集，他们处理速度比较慢。为了很好地解决速度和准确率之间的矛盾，本文提出了基于近邻选择的方式减少字典的数目，极大提高求解速度，同时分类准确率基本没有下降。从各类数据集中，基于近邻选择的模糊重构效果始终优于基于近邻选择的 OPFR，和基于近邻选择的 SRC 的分类效果相当。

4.1 数据集简介

Extended Yale B 数据集^{[43][44]}是由不同的光照条件下的 38 个人的人脸数据组成，包括 2414 张人脸图像，被归一化为 192 x 168。随机选择一半数据作为训练数据(字典)，剩余的数据用于测试。CMU Pose Illumination Expression (PIE)^[45]是由不同光照和表情的 68 个人的人脸数据组成。在本文的实验中，我们选取包含五种姿势(C05, C07, C09, C27, C29)的数据子集。随机选择每个人的 150 张图片作为训练集，剩余的作为测试集。USPS^[46]和 Pendigits^[47]是手写数字数据集，根据 Hull 等提供的意见将 USPS 划分为训练集和测试集，Pendigits 根据 UCI 上提供的

方法划分训练集和测试集。Sonar^[47]也是来自于 UCI 上的数据集，由于该数据集的样本数较少，把数据集进行 10 等分，其中一份用于测试，剩余的作为训练集，进行 10 次交叉验证。各个数据集的简单信息如表 4.1 所示，包含了人脸、手写数字以及其他的数据集。既有高维的大样本的数据集，又有低维的小数据集。

表 4.1 数据集简介

序号	数据集名称	样本数	属性	类别数
1	Extended Yale B	2414	32256	38
2	PIE	11554	4096	64
3	USPS	9298	256	10
4	Pendigits	10992	16	10
5	Sonar	208	60	2

4.2 FR、KFR、KNN 和 FKNN 对比实验

人脸数据的维度较高，本文选取了主成分分析 PCA 进行降维。本节主要对比分析模糊重构 FR 和基于近邻选择的模糊重构 KFR 之间以及 KFR 和 KNN、模糊 KNN(FKNN)之间分类效果的差异。通常情况下，由于 KNN 和 FKNN 中的 K 取 1 至 10，分类效果较好。在本次试验中，KNN 和 FKNN 中的 K 值分别为 3、5、7、9 时，进行了四轮实验，图 4.1 和图 4.2 展示了 KNN 和 FKNN 在四组取值时最好的分类效果。基于近邻选择的模糊重构 KFR 中 K 的选择具有较强的鲁棒性，在本次试验中 K 为固定值，为类别数的 10 倍。用 PCA 将原始数据分别降至 64、96、128、150 维，然后进行分类实验。从图 4.1 中可以看出，FR 和 KFR 的分类准确率相差无几，在 64 维的特征空间中，KFR 的分类准确率要高于 FR。在 128 维的特征空间中，FR 的分类效果优于 KFR。FR 和 KFR 的分类效果始终好于 KNN 和 FKNN，由此可以看出基于 KNN 的快速模糊重构不仅能提高优化速度，有时也能提高准确率。当维度为 64 时，FR 和 KFR 的分类准确率分别是 89.18%和 89.75%，

而 KNN 和 FKNN 的准确率分别为 63.87%和 60.08%，FR 和 KFR 分别相对 KNN 提高了 39.67%和 40.45%，相对 FKNN 分别提高了 40.52%和 49.38%。

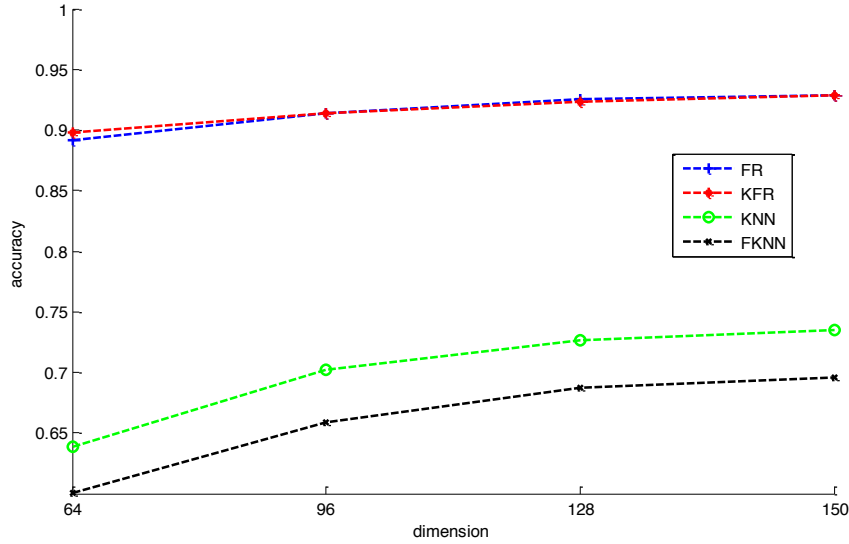


图 4.1 FR、KFR、KNN 以及 FKNN 在 Extend_Yale_B 人脸数据集上的分类准确率对比，横坐标为维数，纵坐标为准确率。

为了更好地分析每类样本的分类情况，图 4.2 展示了四种算法在该数据集的不同类别样本上的分类准确率。可以看出，模糊重构 FR 和基于近邻选择的模糊重构 KFR 的分类准确率在各类样本中始终高于 KNN，两者的变化趋势基本相同。在第 5、13、14、27 和 33 类样本中，FR 和 KFR 的分类准确率为 100%。对于第 10 类样本，模糊重构相对 KNN 提高了 160.00%，为最大增幅。对于第 13 类样本，提高幅度最小为 10.34%。可见，对于各种类型的数据集上，FR 和 KFR 都比 KNN 具有较强的处理能力，但其体现的优势不尽相同。

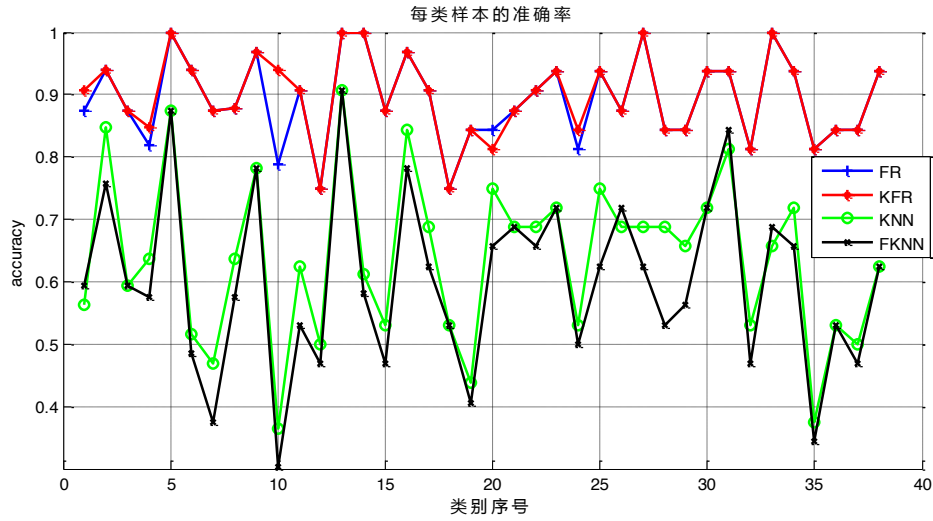


图 4.2 FR、KFR、KNN 以及 FKNN 在不同类别的样本上的分类准确率。其中横坐标表示类别标号，纵坐标表示准确率。

4.3 基于近邻选择的各类算法对比实验

从图 3.1 可以看出，大部分的重构系数都为零，只有部分系数为非零。图 3.2 表明这些参与重构的样本和测试样本属于同一类，也即只有和测试样本较为相似的部分字典参与重构。同时为了提升优化速度，在模糊重构之前先进行近邻选择预处理。在 KNN 中，K 的选择对分类结果较为敏感，合适 K 的取值区间较小。但是基于近邻选择的模糊重构对 K 的选择具有较强的鲁棒性，K 的取值区间较大。选择一个合适的 K，KFR 和 FR 的识别率相差无几，有时甚至 KFR 的分类效果好于 FR，但提高了优化求解的速度。

OPFR 和 SRC 的重构系数都是稀疏的，和模糊重构的情况类似，只有部分样本参与重构，因此只要选择一个合适的 K，基于近邻选择预处理的 OPFR 和 SRC 的分类准确率不会明显下降。图 4.3 表明，在 Yale 人脸数据集上，五种分类算法在不同的维度上的分类准确率随着 K 的变化情况。随着 K 的增大，基于近邻选择的 FR、SRC、OPFR 的分类准确率逐渐增大，当 K 增大到一定程度之后，他们的分类准确率基本不变。KNN 和 FKNN 随着 K 的增大，分类准确率在下降。但 FKNN 下降的幅度和速率较慢。当 K 大于 38(类别数)时，基于近邻选择的 FR、SRC、OPFR 以及 FKNN 的分类准确率趋于一个稳定的值，而 KNN 只有取[1,10]时，分

类准确率较高，其他值对最终分类结果影响较大。

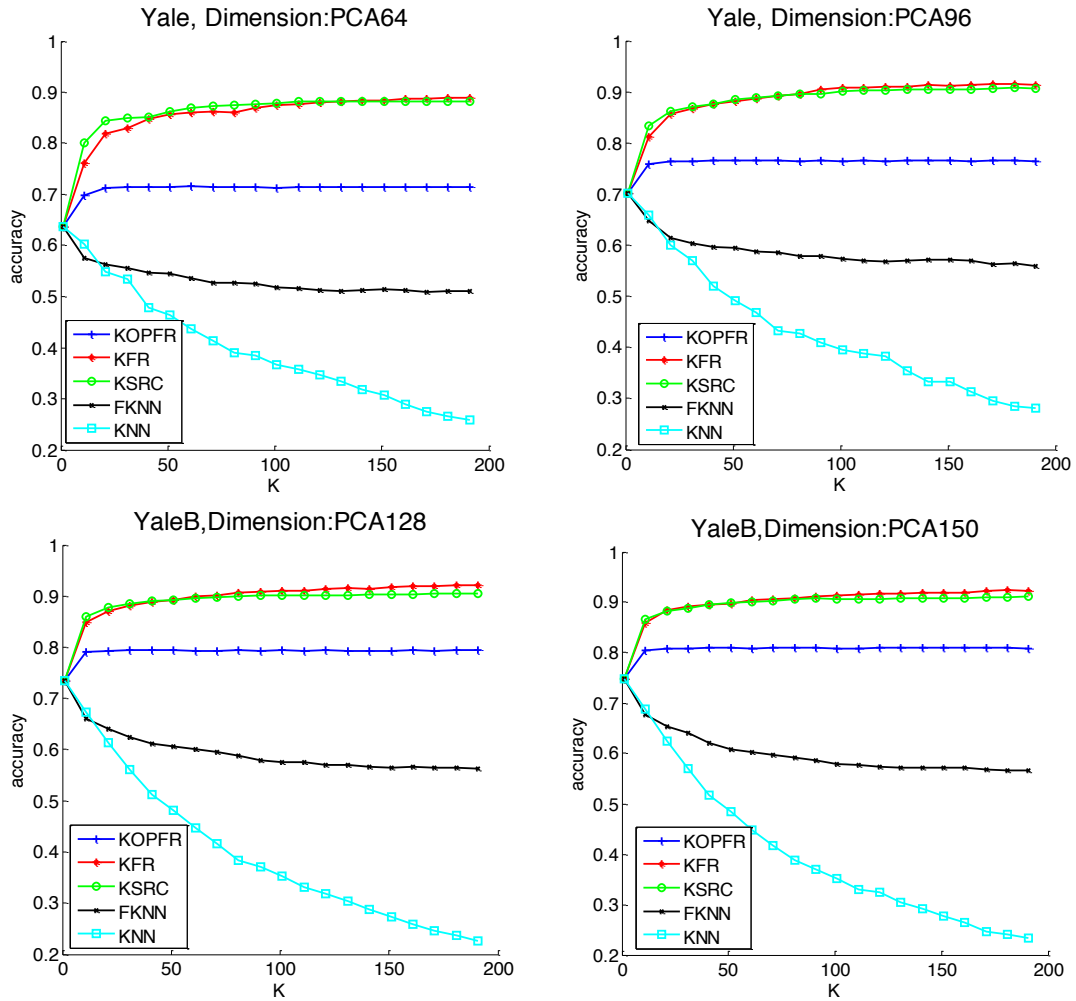


图 4.3 Yale 数据集, 随着 K 取值的变化五种算法的分类效果对比

根据需求选择相应的 K，如果追求准确率，选择较大的 K，可以选择为类别数的 10 倍或训练集的十分之一。如果追求速度，选择较小的 K，可以选择为类别数或训练集的二十分之一。本次所有的实验都是使用 matlab 在 4.00G 内存，3.40GHz 的电脑上运行的，当利用 PCA 降至 128 维时，K=38(类别数)时，分类准确率为 87.76%，运行时间为 56.83s。当 K=380(类别数的 10 倍)时，分类准确率为 91.73%，运行时间为 615.50s。分类准确率提高了 3.97%，而运行时间是原来的 10.83 倍。当利用 LPP(参数设置同 2.4 节)降至 128 维时，K=38 时，分类准确率为 84.53%，运行时间为 63.97s。当 K=380 时，分类准确率为 84.62%，运行时间为

617.85s。分类准确率提高了 0.09%，而运行时间是原来的 9.66 倍。基于近邻选择的 FR、SRC、OPFR 的分类准确率始终比 KNN 和 FKNN 高，基于近邻选择的 FR、SRC 的识别率通常高于 OPFR，他们两者的分类准确率基本相当。

图 4.4 表明，刚开始随着 K 的增大，基于近邻选择的 SRC 的分类效果要优于 FR，但当 K 增大一定程度(类别数的 2 倍)时，两者的分类准确率基本相当。图 4.5 和图 4.6 是在手写数字数据集上的实验结果，基于近邻选择的 FR、SRC、OPFR 三者的分类准确率基本相当。图 4.7 是低维的小数据集 Sonar 上的分类结果，KFR 和 KSRC 的分类准确率基本相当，比 KOPFR 的分类效果好一些。

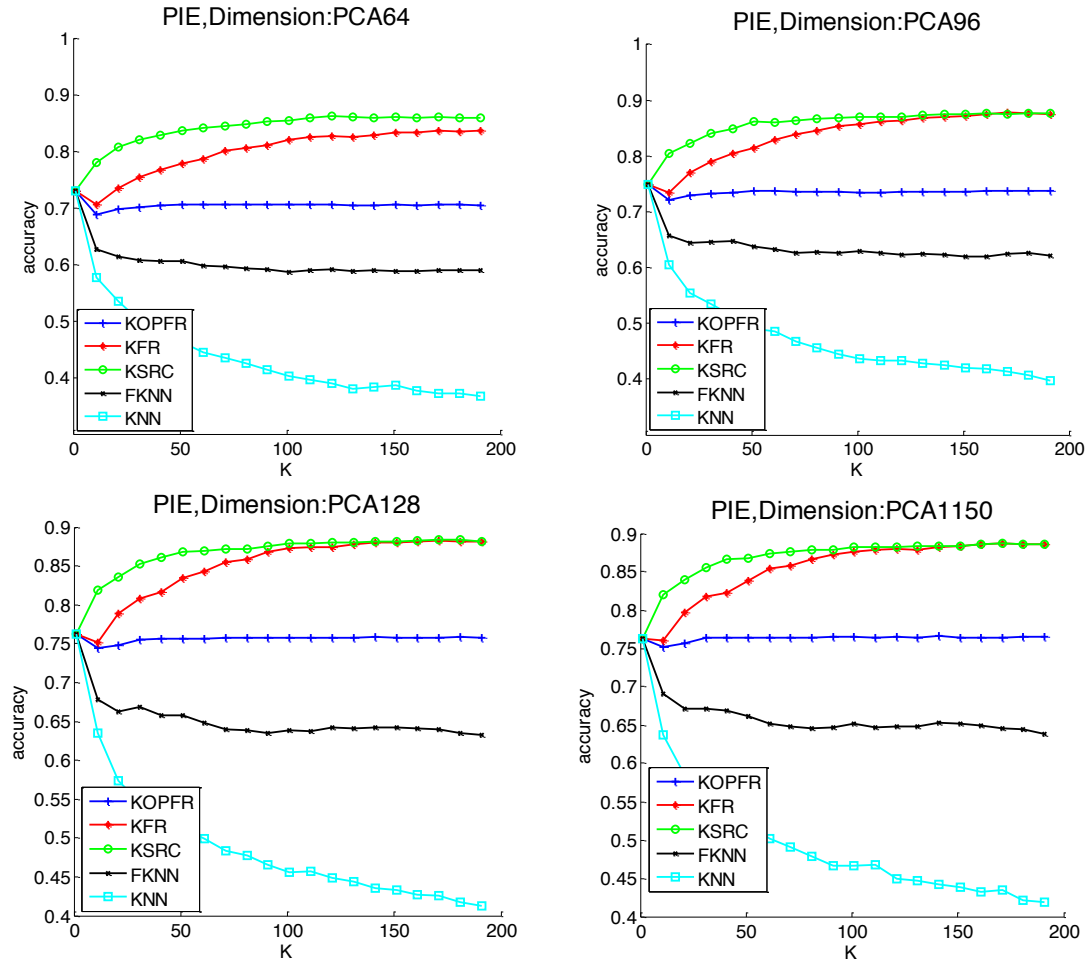
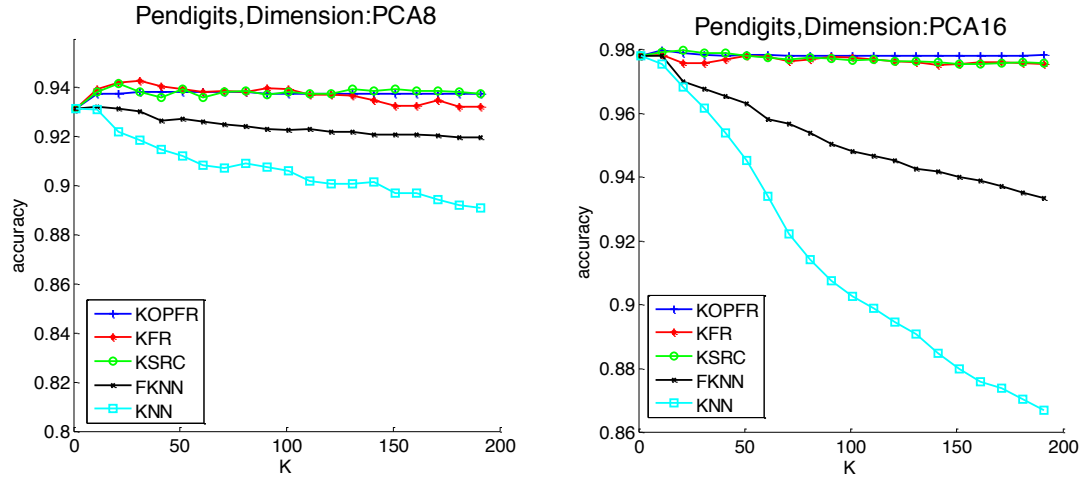
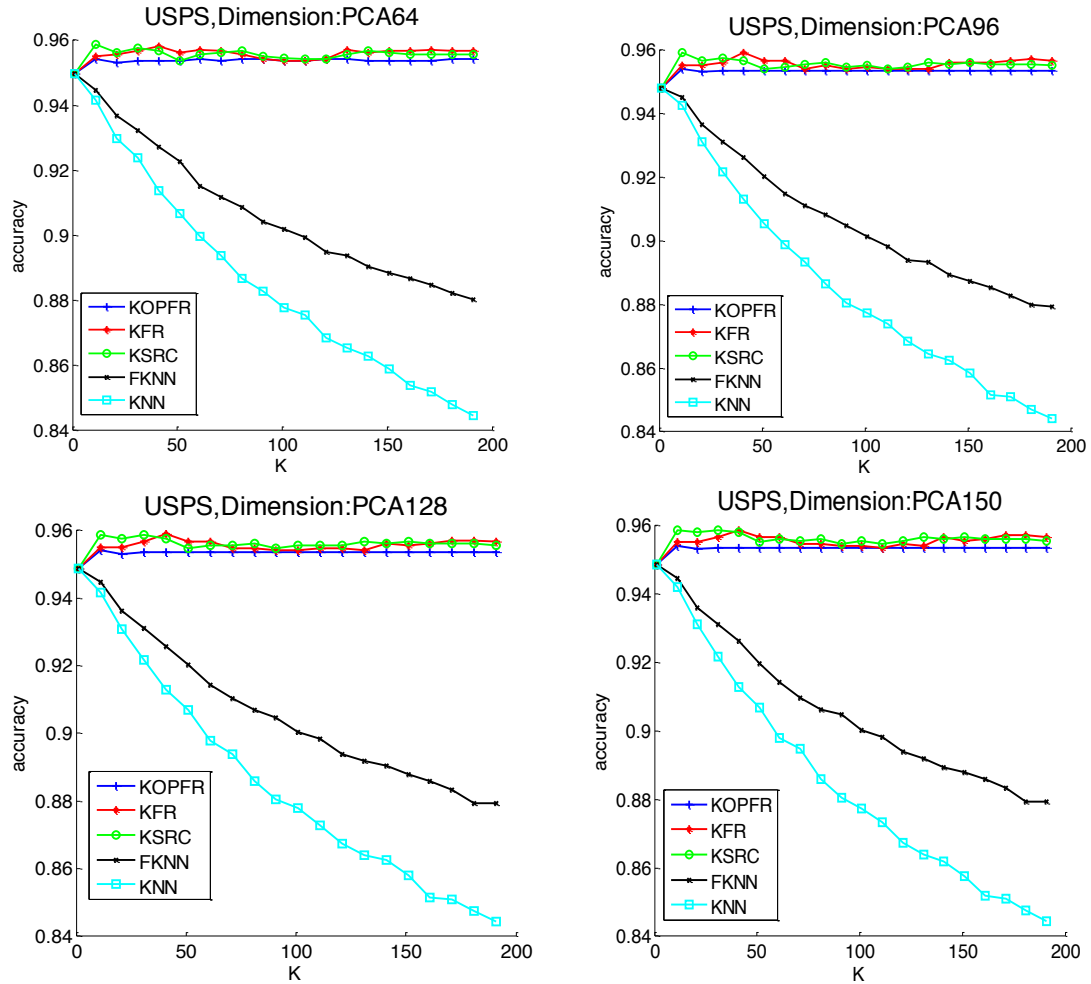


图 4.4 PIE 数据集, 随着 K 取值的变化五种算法的分类效果对比

图 4.5 Pendigits 数据集, 随着 K 取值的变化五种算法的分类效果对比图 4.6 USPS 数据集, 随着 K 取值的变化五种算法的分类效果对比

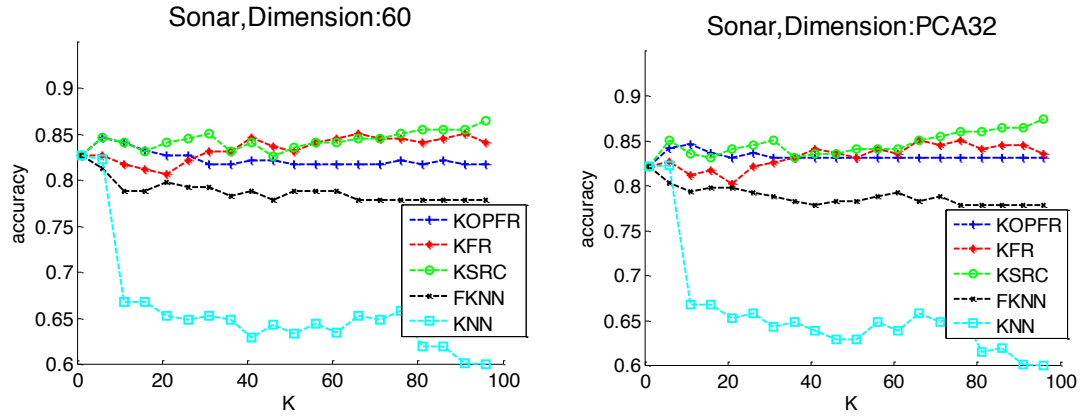


图 4.7 Sonar 数据集, 随着 K 取值的变化五种算法的分类效果对比

本节实验结果表明基于近邻选择的模糊重构对 K 值的选取不敏感, 当 K 增大到一定的数值之后, 分类准确率随着 K 的变化较小。拐点处的 K 值较好地平衡了分类准确率和优化速度。这也是在 2.4 节和 4.2 节中基于近邻选择的模糊重构中 K 选取固定值的原因。

4.4 重构系数分析

在前面的分类实验中, 我们使用最大模糊隶属度原则作为分类判别标准。当最大的隶属度和其他隶属度相差较大时, 如图 3.2 中最大隶属度接近于 1, 其他隶属度接近于 0, 我们可以比较确信地把测试样本归为最大隶属度所属的类。但当它们之间相差比较小时, 此时利用最大隶属度作为分类判别标准就难以让人确信, 测试样本也可能和次大的隶属度所属的类相同。图 4.8 展示了在 Yale 64 维人脸数据集上的实验结果, 有 1088 个测试样本可以根据最大隶属度原则被正确分类, 其中 999 个样本的最大的两个隶属度之比大于 2.0, 89 个小于 2.0。有 47 个样本和次大的隶属度所属的类相同, 其中仅有 5 个最大的两个隶属度之比大于 2.0。有 18 个样本和第三大隶属度所属的类相同, 其中仅有 3 个隶属度之比大于 2.0。由此可见, 大部分测试样本可以根据最大隶属度原则进行分类, 但当最大隶属度和次大隶属度相差较小时, 容易导致错误分类。

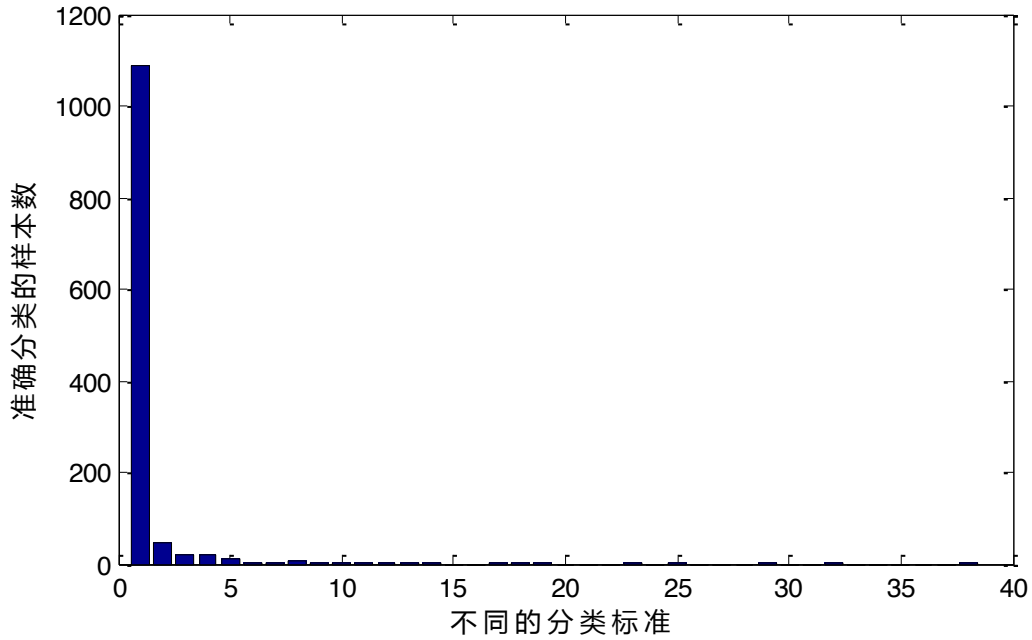


图 4.8 测试样本利用不同的隶属度强度作为分类标准时被正确分类的样本数分布情况。横坐标代表分类准则，从左到右表示隶属度的排序，最左边和最右边分表表示用最大隶属度和最小隶属度作为分类决策标准。纵坐标表示正确分类的数目。

为了进一步分析模糊隶属度的分布情况，图4.9左上角的图显示了在Yale数据集上利用最大隶属度作为分类标准，也即测试样本所属的类和最大的隶属度所属的类相同时，被正确分类的样本的隶属度分布情况，从中可以看出最大隶属度远大于其他隶属度值。图4.9剩余的三个图分别显示了利用次大、第三大隶属度和第四隶属度原则作为分类标准被正确分类的样本的隶属度分布情况，也即这些样本是利用最大隶属度原则被错误分类的，从图中可以看出最大隶属度和和其他隶属度相差无几，此时我们很难以较大的可信度预测测试样本的类标签。通过模糊重构，我们可以判断哪些样本容易区分，哪些样本较易误判。比如我们优化求解得到测试样本相对类1至类5的隶属度分别为0.32、0.24、0.18、0.08、0.06、0.05，而相对类6至类38的隶属度接近于0。此时我们可以比较确信地把该测试样本归到前两类或前三类的中的某一类。如果我们最终只是把测试样本分到它最有可能所在的两个类中的话，模糊重构算法的分类准确率为93.03%。对于监控和医疗诊断领域的分类问题，需要较高的召回率(recall)。该算法可以很好地解决此类问题。

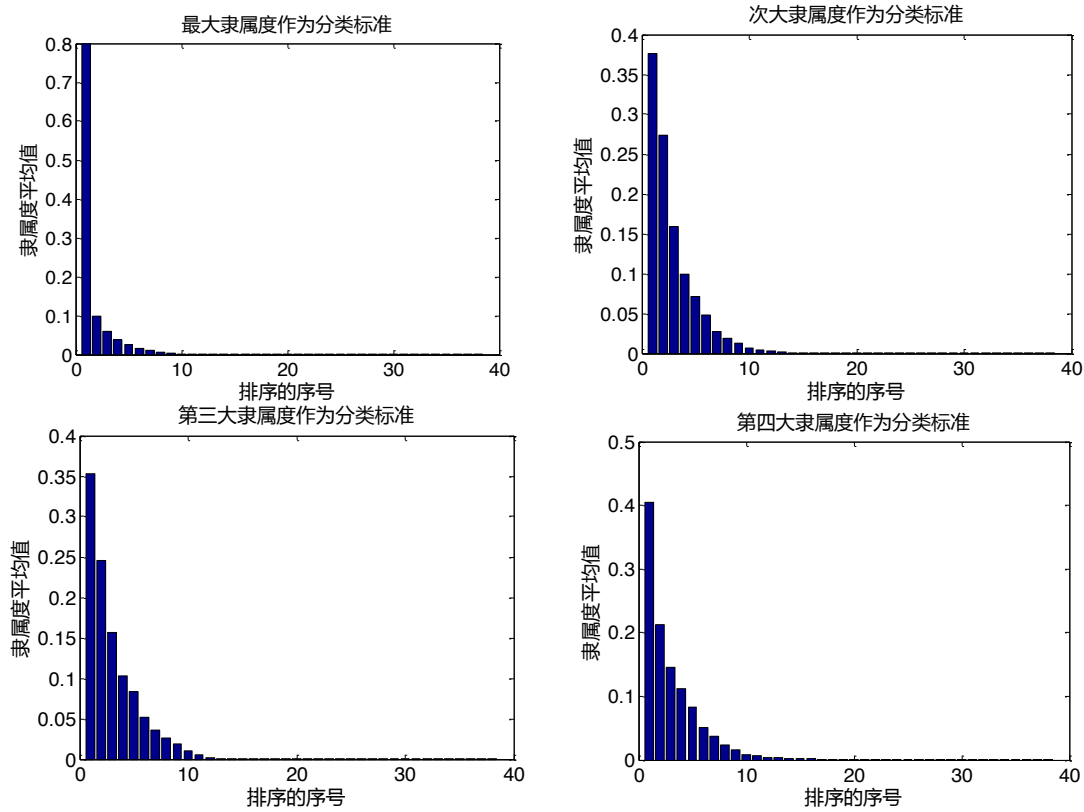


图 4.9 不同的分类标准下,被正确分类的样本相对各类隶属度平均分布情况。横坐标表示所有利用该分类标准被正确分类的测试样本相对各类的隶属度按从大到小排序。纵坐标表示测试样本相对各类(共 38 类)隶属度的统计平均值。

4.5 调整重构系数改善分类性能

由 4.4 节分析可知,当最大的两个隶属度差值不是很大时,我们很难以较大的可信度来判断该样本属于那个类别,但是他们很有可能属于隶属度较大的几个类中的某一个类别。我们可以利用该信息,在系数较大的几类样本中再次重构,进而确定隶属这几类的程度。从图 4.8 可以看出,测试样本所属的类基本在较大的两个隶属度所属的类中。我们可以在这 2 类样本的字典上进行再次重构,确定他们最终属于那个类。当第一次模糊重构时最大的两个隶属度差值不是很大的样本,对它进行二次重构。可以设定当最大的两个隶属度之比小于一个阈值时,我们对该样本进行再次重构。表 4.2 是设定不同的阈值,再次重构调整系数之后的分类准确率。从中可以看出,调整系数可以提高分类准确率,但提高的幅度不是特别大。

表 4.2 选择不同的阈值，调整系数后的分类准确率

Threshold	2.00	2.50
Accuracy (%)	90.00	89.34

由于我们仅仅选择测试样本的隶属度中最大的两类样本作为字典进行重构，减少了字典的规模，此时也限制准确率提高的范围，比如此时准确率最高提高到 93.03%。我们可以根据图 4.8 选择测试样本应该属于隶属度较大的 K 类中的某一类。可以根据统计直方图中的拐点或前 K 类样本占总样本的比例确定 K ，如图 4.8，可以选取 $K=2$ 、3 或 4。

当我们很难以较大的可信度预测一个样本的类标签时，选择与测试样本最可能属于的 K 类训练样本作为字典，进行再次模糊重构，调整重构系数改善分类性能。同样，我们也可以使用其他的算法来改善其分类性能。因为我们已经确定该测试样本最有可能属于 K 类样本中的某一类时，我们可以利用同类样本进行单独的重构，根据重构效果，确定它属于那个类。另外，由于参与重构的样本较少，重构的误差可能较大，而且隶属度较少的训练样本和测试样本的相关性较小，我们可以加入这些隶属度较小的样本作为马甲字典^[48]。可以再次利用前面的模糊重构算法进行重构，根据重构系数进行分类。

通过模糊重构算法或基于近邻选择的模糊重构算法，我们可以正确地对大多数样本进行分类。对于剩余的少数容易误判的样本，可以判定其可能属于几个类。由于从训练中剔除其他无关类样本信息，这样可以大大减少搜索空间。然后可以利用任何模式分类方法(如 SVM，模糊 SVM，神经网络等)进行分类，这样相对于直接在原始训练集上进行学习，可以提高分类的准确率。

4.6 本章小结

本文主要从分类效果方面对比了模糊重构 FR 和基于近邻选择的模糊重构 KFR，当 K 取一个合适的数值时，两者的分类效果基本相当。对比 KFR 和 KNN、FKNN，KFR 的分类准确率始终高于 KNN 和 FKNN。随着 K 的增大，KFR 的分

类准确率逐渐提高,一般 K 等于数据的类别数目时,图像中出现一个拐点,通常情况下, K 取此值较为合适。在拐点之后,随着 K 的增大,分类准确率有小幅度的变化。但对于 KNN,在拐点之后,随着 K 的增大,分类准确率变化较大,且有减少的趋势。可见,基于近邻选择的模糊重构中合适 K 的取值范围较大,KNN 中合适 K 的取值范围较小。若追求高准确率,可以选择一个较大的 K ,如 K 为数据的类别数的 10 倍。若追求高速度,可以选择一个较小的 K ,如 K 为数据的类别数。在人脸、手写数字以及其他的标准数据集上,基于近邻选择的模糊重构,基于近邻选择的稀疏表示的分类效果优于基于近邻选择的保序模糊重构,他们三者的分类准确率始终高于 KNN 和 FKNN。

通过分析模糊重构的系数,我们发现当最大隶属度明显大于次大隶属度时,以比较大的可信度预测该样本的类标签,这些样本易于区分。当最大隶属度和次大隶属度相差无几时,很难确定该样本属于哪个类别,这些样本中可能包含较大的噪声。当最大隶属度和次大隶属度相当时,我们尝试着调整其系数,更好地区分这些样本。

第5章 总结和展望

5.1 全文总结

模糊隶属函数是模糊集理论的重要概念，可以用于表示信息的不确定性和对象的语义信息。现在处于信息时代，我们希望挖掘数据中潜在有价值的信息，从数据中学习隶属函数。为此，我们提出了模糊重构算法。

在字典上进行模糊重构，本文中字典是由训练数据组成，一条字典相当于一条模糊规则，重构系数可以看作模糊规则中的权重。而且此时模糊规则的可解释性较强，易于选取。但模糊分类系统难以处理高维数据，模糊规则的选取较为困难。通过模糊重构，根据重构系数获得模糊隶属函数，根据最大隶属度原则进行分类决策。同时也可以得到一个重构结果。通过分析重构系数，我们可以确定哪些样本容易区分，哪些样本较难判别。对于较难判别的样本，可以通过调整系数等方式改善分类性能。

由于优化时，字典的规模较大，重构时间较长。通常情况下，只有部分字典参与重构，可以选择和测试样本较为相似的部分训练样本作为重构时的字典，进而可以大大提高优化速度。KNN 和 FKNN 中对 K 的选取较为敏感，但基于近邻选择的模糊重构中 K 的取值范围较大。如果追求准确率，可以选择一个较大的 K 。如果追求速度，我们可以选取一个较小的 K 。基于近邻选择的模糊重构较好地平衡了优化速度和分类准确率之间的矛盾。基于近邻选择的模糊重构可以看作是一般化的 KNN 算法，修改了 KNN 中判别规则。当重构系数都相等时，基于近邻选择的模糊重构就退变为 KNN。在人脸、手写数字以及其他的标准数据集上，KFR 的分类准确率始终高于 KNN 和 FKNN，基于近邻选择的模糊重构，基于近邻选择的稀疏表示分类效果优于基于近邻选择的保序模糊重构，前两者的分类效果基本相当。

在模糊重构之前，需要进行数据降维预处理，我们对比了五种无监督的(PCA, NMF, Isomap, LLE, LPP)和一种有监督(LDA)的数据降维方法。数据降维方法也是机器学习、数据挖掘、计算机视觉等领域中常用的数据预处理方法。从数据

可视化和分类准确率两个方面评价了数据降维方法的性能。模糊重构对降维方法具有较强的鲁棒性。

5.2 未来的研究工作展望

模糊重构算法是一项基础性的研究算法，在很多方面具有广泛的应用。本文主要从分类的角度对比了其实验效果，在字典学习、聚类问题和回归等方面具有较大的潜在应用价值。下面从这几个方面分别阐述。

(1) 字典学习。

在前面的模糊重构及其相关的算法中，字典是由训练样本组成。设计具有区分性的字典，模糊重构算法可以取得更好的分类效果。可以基于迭代的思想学习字典。首先基于已有的字典进行模糊重构，然后固定重构系数更新字典。反复地迭代学习具有区分性的字典。这可以看做为一般化的 k -means 算法，当字典只有 K 个，模糊重构系数只有一个为非零，其他所有的值都为零时，此算法就变化了 k -means 算法。由于在重构的过程中限制系数在 0 至 1 之间，可能会像非负矩阵分解 NMF 一样学习一些特殊的字典，例如部分脸。

(2) 聚类问题。

聚类是无监督学习，也是机器学习和数据挖掘中的一个重要问题，。基于稀疏表示的聚类方法(SSC)^[49]根据稀疏表示构建相似性矩阵。然后利用谱聚类进行数据的聚类。这种算法可以较好地处理包含噪声、数据缺失等情况下的聚类问题。SSSC^[50]是基于 SSC 的聚类算法，较好地解决了 SSC 的可扩展性和 out-of-sample 问题。SSSC 首先进行抽样，利用 SSC 算法进行聚类。对剩余的样本进行稀疏编码，由于 ℓ_1 -norm 优化时间较长，为了加快求解速度，此时采用了 ℓ_2 -norm 的形式，具有解析解。最后根据最小残差法进行分类。利用模糊重构同样可以构建对象的相似性矩阵，然后可以利用该相似性矩阵进行谱聚类或流行降维，类似于 LPP。

(3) 回归问题。

回归是有监督学习,也是机器学习中的一个重要问题。在实际生活中,有很多的应用场景。如股价的预测和市场规趋势的预测等。利用模糊重构可以计算一个样本和字典之间的关系,根据重构的系数作为权重,样本的输出等于字典输出的加权之和。基于近邻选择的模糊重构用于解决回归问题时,可以看作 KNN 回归的一般化情况。

(4) 保序关系。

在保序模糊重构时,我们利用欧式距离定义他们之间的序关系。但在高维中,第二章曾介绍过利用欧式距离度量两个样本之间的相似性是有问题的。可以探索利用其它方式定义序关系。如测地距离和余弦相似性等。也可以根据实际的情况,由用户自定义一些序关系。同时所有样本上的保序关系是比较强的约束,可以放松这种约束关系,在类内保序,类间不保序。通过初步的实验分析,在分类问题中,按类保序的模糊重构的分类效果优于在所有样本上的保序模糊重构。

前面我们分别从字典学习、聚类问题、回归问题探讨了模糊重构潜在的应用。在解决问题时,需要根据问题的性质调整算法,这样才能够取得更好的效果。

参考文献

- [1] Zadeh L A. Fuzzy sets[J]. Information and control, 1965, 8(3): 338-353.
- [2] Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: A strategy employed by V1?[J]. Vision research, 1997, 37(23): 3311-3325.
- [3] 李航. 统计学习方法[M]. 2012.
- [4] Medasani S, Kim J, Krishnapuram R. An overview of membership function generation techniques for pattern recognition[J]. International Journal of approximate reasoning, 1998, 19(3): 391-417.
- [5] Ananthi V P, Balasubramaniam P, Lim C P. Segmentation of gray scale image based on intuitionistic fuzzy sets constructed from several membership functions[J]. Pattern Recognition, 2014, 47(12): 3870-3880.
- [6] Hanmandlu M, Murthy O V R. Fuzzy model based recognition of handwritten numerals[J]. Pattern Recognition, 2007, 40(6): 1840-1854.
- [7] Lei Z, Ren-Hou L. Designing of classifiers based on immune principles and fuzzy rules[J]. Information Sciences, 2008, 178(7): 1836-1847
- [8] Ishibuchi H, Nozaki K, Yamamoto N, et al. Selecting fuzzy if-then rules for classification problems using genetic algorithms[J]. Fuzzy Systems, IEEE Transactions on, 1995, 3(3): 260-270.
- [9] Cordon O, Herrera F, Villar P. Analysis and guidelines to obtain a good uniform fuzzy partition granularity for fuzzy rule-based systems using simulated annealing[J]. International Journal of Approximate Reasoning, 2000, 25(3): 187-215.
- [10] Ishibuchi H, Nozaki K, Yamamoto N, et al. Selecting fuzzy if-then rules for classification problems using genetic algorithms[J]. Fuzzy Systems, IEEE Transactions on, 1995, 3(3): 260-270.
- [11] Sanz J A, Fernández A, Bustince H, et al. IVTURS: A Linguistic Fuzzy Rule-Based Classification System Based On a New Interval-Valued Fuzzy

- Reasoning Method With Tuning and Rule Selection[J]. IEEE T. Fuzzy Systems, 2013, 21(3): 399-411.
- [12] Wu X, Kumar V, Quinlan J R, et al. Top 10 algorithms in data mining[J]. Knowledge and Information Systems, 2008, 14(1): 1-37.
- [13] Cover T M, Hart P E. Nearest neighbor pattern classification[J]. Information Theory, IEEE Transactions on, 1967, 13(1): 21-27.
- [14] 苏毅娟,邓振云,程德波,宗鸣. 大数据下的快速KNN分类算法[J]. 计算机应用研究, 2016, 04:1-6.
- [15] Keller J M, Gray M R, Givens J A. A fuzzy k-nearest neighbor algorithm[J]. Systems, Man and Cybernetics, IEEE Transactions on, 1985 (4): 580-585.
- [16] Derrac J, García S, Herrera F. Fuzzy nearest neighbor algorithms: Taxonomy, experimental analysis and prospects[J]. Information Sciences, 2014, 260: 98-119.
- [17] Chen H L, Huang C C, Yu X G, et al. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach[J]. Expert systems with applications, 2013, 40(1): 263-271.
- [18] Olshausen B A. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J]. Nature, 1996, 381(6583): 607-609.
- [19] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009, 31(2): 210-227.
- [20] Tang Y, Lee Z. How to adjust the distribution of nonzero elements in sparse representation: A granular locality-preserving approach[J]. Information Sciences, 2014, 289: 25-40.
- [21] Naseem I, Togneri R, Bennamoun M. Linear regression for face recognition[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010, 32(11): 2106-2112.
- [22] Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: Which helps face recognition?[C]. Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 471-478.

- [23] Zhang H, Wang F, Chen Y, et al. Sample pair based sparse representation classification for face recognition[J]. Expert Systems with Applications, 2016, 45: 352-358.
- [24] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches[C]. Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.
- [25] Mei X, Ling H. Robust visual tracking and vehicle classification via sparse representation[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2011, 33(11): 2259-2272.
- [26] Liu H, Liu Y, Sun F. Traffic sign recognition using group sparse coding[J]. Information Sciences, 2014, 266: 75-89.
- [27] Aharon M, Elad M, Bruckstein A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation[J]. Signal Processing, IEEE Transactions on, 2006, 54(11): 4311-4322.
- [28] Ma L, Wang C, Xiao B, et al. Sparse representation for face recognition based on discriminative low-rank dictionary learning[C]. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 2586-2593.
- [29] Yang M, Dai D, Shen L, et al. Latent Dictionary Learning for Sparse Representation based Classification[C]. Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014: 4138-4145.
- [30] Jolliffe I. Principal component analysis[M]. John Wiley & Sons, Ltd, 2002.
- [31] Bishop C M. Pattern recognition and machine learning[M]. springer, 2006.
- [32] Schölkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural computation, 1998, 10(5): 1299-1319.
- [33] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788-791.
- [34] Lee D, Seung H S. Algorithms for non-negative matrix factorization[C]. Advances in neural information processing systems. 2001: 556-562.
- [35] Tenenbaum J B, De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2323.

- [36] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.
- [37] He X, Niyogi P. Locality Preserving Projections[C]. Advances in Neural Information Processing Systems. 2004: 153-160.
- [38] Duda R O, Hart P E, Stork D G. 模式分类[M]. 机械工业出版社, 2003.
- [39] Donoho D L. For most large underdetermined systems of linear equations the minimal ℓ_1 - norm solution is also the sparsest solution[J]. Communications on pure and applied mathematics, 2006, 59(6): 797-829.
- [40] Kim S J, Koh K, Lustig M, et al. An interior-point method for large-scale ℓ_1 -regularized least squares[J]. Selected Topics in Signal Processing, IEEE Journal of, 2007, 1(4): 606-617.
- [41] Mairal J, Bach F, Ponce J, et al. Online learning for matrix factorization and sparse coding[J]. The Journal of Machine Learning Research, 2010, 11: 19-60.
- [42] Y. Tang, Z. Lee, Ordering preserving fuzzy reconstruction as a representation for semantics of linguistic labels[J].
- [43] Georgiades A S, Belhumeur P N, Kriegman D. From few to many: Illumination cone models for face recognition under variable lighting and pose[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001, 23(6): 643-660.
- [44] Lee K C, Ho J, Kriegman D. Acquiring linear subspaces for face recognition under variable lighting[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2005, 27(5): 684-698.
- [45] Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression (PIE) database[C]. Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on. IEEE, 2002: 46-51.
- [46] Hull J J. A database for handwritten text recognition research[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1994, 16(5): 550-554.
- [47] K. Bache, M. Lichman, UCI Machine Learning Repository, 2015.
- [48] Deng W, Hu J, Guo J. Extended SRC: Undersampled face recognition via intraclass variant dictionary[J]. Pattern Analysis and Machine Intelligence, IEEE

- Transactions on, 2012, 34(9): 1864-1870.
- [49] Elhamifar E, Vidal R. Sparse subspace clustering[C]. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 2790-2797.
- [50] Peng X, Zhang L, Yi Z. Scalable sparse subspace clustering[C]. Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013: 430-437.

致谢

光阴似箭，研究生阶段的学习生涯就要匆匆结束了。大三暑假第一次来到浙江大学计算机学院实习，老师们深厚的知识积累和宽广的视野、师兄师姐拼搏进取的精神、实验室浓厚的学术氛围等深深地吸引了我。在即将毕业之时，此情此景仍然历历在目，非常感谢孙凌云老师当时对我的指导。

在这里，首先要感谢我的导师汤老师。本文是在汤老师的指导下完成的，在理论分析和实验的过程中，遇到了很多的问题，汤老师总会抽出时间与我讨论，并不厌其烦地为我分析讲解，让我对自己的研究领域有了深刻的认识。汤老师严谨的治学态度和对科研的极大热情深深地影响了我。他提出了大的研究方向和长期的研究计划，对自己所提出的理论满怀热情并投入了大量的精力进行探索，制定了短期的研究计划充实和丰富所提出的理论。从汤老师身上我不仅学习了无限的专业知识，更学会了很多做人做事的道理，对以后的人生发展有很大的指导作用。

另外要感谢实验室的师兄师姐师弟师妹对我完成论文中提供了一些有意义的建议和意见，与他们的交流中扩展了我的知识范围，提高了我对专业知识的认识。感谢很多小伙伴们在研究生期间对我的帮助，感谢你们的陪伴，丰富了校园生活，给我留下了美好的回忆。

最后，感谢我的父母和占永杰同学对我无条件的包容和支持，是你们让我更好地完成了学业。

代成超

2016 年 3 月